

new 4 data
02/98

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 15:13:38 ; Search time 10.24 Seconds
(without alignments)
247.549 Million cell updates/sec

Title: US-09-367-009-3
Perfect score: 392
Sequence: 1 DSCCKLLEDVETINSIDIS.....TLKNFGIMMHTVYDSIWCNL 74

Gapop 10.0 , Gapext 0.5

93435 seqs, 34255486 residues

93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Match	Query Length	ID	Description
1	390	99.5	95	1 MGBB_HUMAN	O75556 homo sapien
2	208	53.1	93	1 MGBA_HUMAN	Q13296 homo sapien
3	163	41.6	95	1 PSC3_RAT	P02780 rattus norv
4	70	17.9	303	1 FTSY_RICPR	O05948 rickettsia
5	66.5	17.0	234	1 RADG_HAEIN	P44952 haemophilus
6	65.5	16.7	131	1 YRS9_MYCTU	O33301 mycobacteri
7	65.5	16.7	603	1 VEL_HPV08	P06420 human papil
8	64.5	16.5	603	1 VEL_HPV21	P05759 human papil
9	64.5	16.5	611	1 DYNL_ARATH	P42697 arabidopsis
10	61.5	15.7	162	1 IL15_MOUSE	P48346 mus musculu
11	61.5	15.7	570	1 HEM1_KLULA	P78698 kluyveromyc
12	61.5	15.7	734	1 METE_THEMA	Q9x112 thermotoga
13	61	15.6	605	1 VEL_HPV14	P36721 human papil
14	61	15.6	663	1 TAZ1_SCHPO	P79005 schizosacch
15	61	15.6	732	1 KUB6_MOUSE	P27841 mus musculu
16	60.5	15.4	430	1 SERC_ARATH	Q96255 arabidopsis
17	60.5	15.4	684	1 TC10_YEAST	P50273 saccharomyc
18	60	15.3	268	1 TRPC_ACICA	P00911 acinetobact
19	60	15.3	606	1 VEL_HPV5B	P26542 human papil
20	60	15.3	790	1 V90K_AMVLE	P03593 alfalfa mos
21	60	15.3	1071	1 PR16_YEAST	P15938 saccharomyc
22	59.5	15.2	162	1 IL15_CERAE	P40221 cercopithe
23	59.5	15.2	162	1 IL15_HUMAN	P40933 homo sapien
24	59.5	15.2	162	1 IL15_MACMU	P48092 macaca mula
25	59.5	15.2	925	1 YE9G_SCHPO	O13776 schizosacch
26	59.5	15.2	1177	1 F307_MYCGE	P47549 mycoplasma
27	59	15.1	604	1 VEL_HPV12	Q05134 human papil
28	59	15.1	606	1 VEL_HPV05	P06920 human papil
29	59	15.1	690	1 Y023_NPVAC	P41428 autographa
30	59	15.1	2144	1 GLT1_YEAST	O12680 saccharomyc
31	59	15.1	4540	1 DYHC_PARTE	O27171 paramecium
32	58.5	14.9	322	1 DIDH_RAT	P23457 rattus norv
33	58.5	14.9	606	1 RA17_SCHPO	P50531 schizosacch

34	58	14.8	113	1	P13_MYCMY	O05290 mycoplasma
35	58	14.8	1128	1	DNBI_HSVSA	P24910 herpesvirus
36	57.5	14.7	162	1	IL15_PIG	Q95253 sus scrofa
37	57.5	14.7	300	1	MOVFP_AMVLE	P03595 alfalfa mos
38	57.5	14.7	359	1	PD12_SCHPO	O13811 schizosacch
39	57.5	14.7	527	1	G19P_HUMAN	P14314 homo sapien
40	57.5	14.7	756	1	METE_BACHD	O9kfl1 bacillus ha
41	57.5	14.7	871	1	SC10_YEAST	Q06245 saccharomyc
42	57	14.5	102	1	BOLA_VIBAL	Q56585 vibrio algi
43	57	14.5	281	1	E2F6_HUMAN	O75461 homo sapien
44	57	14.5	299	1	HIS1_ECOLI	P10366 escherichia
45	57	14.5	604	1	VEL_HPV36	P50808 human papil

ALIGNMENTS

RESULT 1
MGBB_HUMAN
ID MGBB_HUMAN STANDARD; PRT; 95 AA.
AC O75556;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MAMMAGLOBIN B PRECURSOR (MAMMAGLOBIN 2) (LIPOPHILIN C) (LACRYGLOBIN).
GN MGB2 OR UGB3 OR LIPHC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99026127; PubMed=9806831;
RA Becker R.M., Darrow C., Zimonjic D.B., Popescu N.C., Watson M.A.,
RA Fleming T.P.;
RT "Identification of mammaglobin B, a novel member of the uteroglobin
RT gene family";
RL Genomics 54:70-78(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99167354; PubMed=10066439;
RA Zhao C., Nguyen T., Yusifov T., Glasgow B.J., Lehrer R.I.;
RT "Lipophilins: human peptides homologous to rat prostatein.";
RL Biochem. Biophys. Res. Commun. 256:147-155(1999).
RN [3]
RP SEQUENCE OF 19-85.
RX TISSUE=Tears;
RX MEDLINE=98163342; PubMed=9504814;
RA Molloy M.P., Bollis S., Herbert B.R., Ou K., Tyler M.I., van Dyk D.D.,
RA Willcox M.D., Gooley A.A., Williams K.L., Morris C.A., Walsh B.J.;
RT "Establishment of the human reflex tear two-dimensional polyacrylamide
RT gel electrophoresis reference map: new proteins of potential
RT diagnostic value";
RL Electrophoresis 18:2811-2815(1997).
RN [4]
RP SEQUENCE OF 19-48 AND 60-78, AND MASS SPECTROMETRY.
RX MEDLINE=98385871; PubMed=9720917;
RA Lehrer R.I., Xu G., Abduragimov A., Dinh N.N., Qu X.-D., Martin D.,
RA Glasgow B.J.;
RT "Lipophilin, a novel heterodimeric protein of human tears";
RL FEBS Lett. 432:163-167(1998).
CC -I- FUNCTION: MAY BIND ANDROGENS AND OTHER STEROIDS, MAY ALSO BIND
CC ESTRADIOL, A CHEMOTHERAPEUTIC AGENT USED FOR PROSTATE CANCER.
CC MAY BE UNDER TRANSCRIPTIONAL REGULATION OF STEROID HORMONES.
CC -I- SUBUNIT: HETERODIMER OF A LIPOPHILIN A AND A LIPOPHILIN C
CC (MAMMAGLOBIN B) MONOMER ASSOCIATED WITH HEAD.
CC -I- TISSUE SPECIFICITY: EXPRESSED IN THYMUS, TRACHEA, KIDNEY, STEROID
CC RESPONSIVE TISSUES (PROSTATE, TESTIS, UTERUS, BREAST AND OVARY),
CC AND SALIVARY GLAND.
CC -I- MASS SPECTROMETRY: MH=8854.94; METHOD=ELECTROSPRAY; RANGE=19-95.
CC -I- SIMILARITY: BELONGS TO THE UTEROGLOBIN FAMILY. LIPOPHILIN
CC SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; AF071219; AAC79996.1; -
DR EMBL; AT224173; CAAL1865.1; -
DR MIM; 604398; -
DR InterPro; IPR000329; -
DR Pfam; PF01099; Uteroglobin; 1.
DR PROSITE; PS00403; UTEROGLOBIN_1; FALSE_NEG.
DR PROSITE; PS00404; UTEROGLOBIN_2; FALSE_NEG.
DR Signal; Glycoprotein.
DR SIGNAL; 1 18
DR CHAIN 19 95 MAMMAGLOBIN B.
DR CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
DR SEQUENCE 95 AA; 10884 MW; 0719738289F89F8D CRC64;

Query Match 99.5%; Score 390; DB 1; Length 95;
Best Local Similarity 98.6%; Pred. No. 1.4e-35;
Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

19 DSGCKLEDMVKETNSDISIPEYKELQEFIDSDAAAEAMGKFCFLNQSHRTLNKFG 60
61 LMHTVYDSIWCNL 74
79 LMHTVYDSIWCNM 92

RESULT 2
ID MG8A_HUMAN STANDARD; PRT; 93 AA.
Q13296;
01-NOV-1997 (Rel. 35, Created)
01-NOV-1997 (Rel. 35, Last sequence update)
01-OCT-2000 (Rel. 40, Last annotation update)
MAMMAGLOBIN A PRECURSOR (MAMMAGLOBIN 1).
MG8A OR UGB2.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
TISSUE=Breast;
MEDLINE=96223698; PubMed=8631025;
Watson M.A., Fleming T.P.;
"Mammaglobin, a mammary-specific member of the uteroglobin gene
family, is overexpressed in human breast cancer.";
Cancer Res. 56:860-865(1996).
[2]
SEQUENCE FROM N.A.
MEDLINE=98147371; PubMed=9488047;
Watson M.A., Darlow C., Zimonjic D.B., Popescu N.C., Fleming T.;
"Structure and transcriptional regulation of the human mammaglobin
gene, a breast cancer associated member of the uteroglobin gene
family localized to chromosome 11q13.";
Oncogene 16:817-824(1998).
-1- TISSUE SPECIFICITY: MAMMARY-SPECIFIC. OVER-EXPRESSED IN BREAST
CANCER.
-1- SIMILARITY: BELONGS TO THE UTEROGLOBIN FAMILY. LIPOPHILIN
SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U33147; AAC50375.1; -
DR EMBL; AF015224; AAC39608.1; -
DR MIM; 605562; -
DR InterPro; IPR000329; -
DR Pfam; PF01099; Uteroglobin; 1.
DR PROSITE; PS00403; UTEROGLOBIN_1; FALSE_NEG.
DR PROSITE; PS00404; UTEROGLOBIN_2; 1.
DR Signal; Glycoprotein.
DR SIGNAL; 1 18 POTENTIAL.
DR CHAIN 19 93 MAMMAGLOBIN A.
DR CARBOHYD 53 53 N-LINKED (GLCNAC. .) (POTENTIAL).
DR CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
DR SEQUENCE 93 AA; 10499 MW; 2896E8C43BF053E2 CRC64;

Query Match 53.1%; Score 208; DB 1; Length 93;
Best Local Similarity 53.4%; Pred. No. 5.9e-16;
Matches 39; Conservative 12; Mismatches 22; Indels 0; Gaps 0;

QY 2 SGCILLEDMVKETNSDISIPEYKELQEFIDSDAAAEAMGKFCFLNQSHRTLNKFG 61
Db 20 SGCPLLENVSKTINPQVSKTEYKELQEFIDSDAAAEAMGKFCFLNQSHRTLNKFG 79
QY 62 MMHTVYDSIWCNL 74
Db 80 FMQIYDSSLCDL 92

RESULT 3
ID PSC3_RAT STANDARD; PRT; 95 AA.
AC P02780; O63463;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE PROSTATIC STEROID-BINDING PROTEIN C3 CHAIN PRECURSOR (PROSTATEIN
DE PEPTIDE C3).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83082848; PubMed=6294095;
RA Parker M.G., White R., Hurst H., Needham M., Tilly R.;
RT "Prostatic steroid-binding protein. Isolation and characterization of
RT C3 genes.";
RL J. Biol. Chem. 258:12-15(1983).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=83238526; PubMed=6190812;
RA Viskochil D.H., Perry S.T., Lea O.A., Stafford D.W., Wilson E.M.,
RA French F.S.;
RT "Isolation of two genomic sequences encoding the Mr = 14,000 subunit
RT of rat prostatein.";
RL J. Biol. Chem. 258:8861-8866(1983).
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92165796; PubMed=1537831;
RA Tan J.A., Marschke K.B., Ho K.-C., Perry S.T., Wilson E.M.,
RA French F.S.;
RT "Response elements of the androgen-regulated C3 gene.";
RL J. Biol. Chem. 267:4456-4466(1992).
RN [4]
RP ERRATUM.
RX MEDLINE=92218467; PubMed=1339454;
RA Tan J.A., Marschke K.B., Ho K.-C., Perry S.T., Wilson E.M.,
RA French F.S.;
RL J. Biol. Chem. 267:7958-7958(1992).

NCBI_TaxID=782;

[1]
SEQUENCE FROM N.A.
STRAIN-MADRID E;
MEDLINE=97419517; PubMed=9274032;
Andersson J.O., Andersson S.G.E.;
"Genomic rearrangements during evolution of the obligate
intracellular parasite Rickettsia prowazekii as inferred from an
analysis of 52015 bp nucleotide sequence.";
Microbiology 143:2783-2795(1997).

[2]
SEQUENCE FROM N.A.
STRAIN-MADRID E;
MEDLINE=99039499; PubMed=9823893;
Andersson S.G.E., Zomorodipour A., Andersson J.O.,
Sicherits-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
Eriksson A.-S., Winkler H.H., Kurland C.G.;
"The genome sequence of Rickettsia prowazekii and the origin of
mitochondria.";
Nature 396:133-140(1998).

[3]
SEQUENCE OF 168-303 FROM N.A.
STRAIN-MADRID E, AND B;
MEDLINE=99416441; PubMed=10486973;
Andersson J.O., Andersson S.G.E.;
"Genome degradation is an ongoing process in Rickettsia.";
Mol. Biol. Evol. 16:1178-1191(1999).

-1- FUNCTION: FUNCTIONAL HOMOLOG OF SRP RECEPTOR. PROBABLY INVOLVED IN
THE RECEPTION AND INSERTION OF A SUBSET OF PROTEINS AT THE
MEMBRANE (BY SIMILARITY).

-1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED (BY SIMILARITY).

-1- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

EMBL; Y11784; CAAT2477.1; -.
EMBL; AJ235273; CAAT15202.1; -.
EMBL; AJ238755; CAB56084.1; -.
EMBL; AJ238756; CAB56088.1; -.
InterPro; IPR000897; -.
Pfam; PF00448; SRP54; 1.
PROSITE; PS00300; SRP54; 1.
Signal recognition particle; GTP-binding; RNA-binding; Membrane;
Cell division.
KW NP_BIND 108 115 GTP (BY SIMILARITY).
FT NP_BIND 190 194 GTP (BY SIMILARITY).
FT NP_BIND 254 257 GTP (BY SIMILARITY).
SO SEQUENCE 303 AA; 33313 MW; BF3F7EB9383B7007 CRC64;

MEDLINE-99416441; PubMed-10486973;
 Andersson J.O., Andersson S.G.E.;
 "Genome degradation is an ongoing process in Rickettsia.";
 MOL. BIOL. EVOL. 16:1178-1191(1999).
 -1- FUNCTION: FUNCTIONAL HOMOLOG OF SRP RECEPTOR. PROBABLY INVOLVED IN
 THE RECEPTION AND INSERTION OF A SUBSET OF PROTEINS AT THE
 MEMBRANE (BY SIMILARITY).
 -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED (BY SIMILARITY).
 -1- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.

 THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).

 -|- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR	PROB1B; P50030; SARJ4; 1.	Signal recognition particle; GTP-binding; RNA-binding; Membrane;
FT	Cell division.	
FT	NP_BIND 108	GTP (BY SIMILARITY).
FT	NP_BIND 190	GTP (BY SIMILARITY).
FT	NP_BIND 254	GTP (BY SIMILARITY).
SO	SEQUENCE 303 AA; 33313 MW; BF3F7EB933B7007 CRC64;	

FT	254	GTP (BY SIMILARITY).
NP_BIND	250	
SEQUENCE	303 AA;	33313 MW; BF3F7EB9383B7007 CRC64;

Query Match. 17.9%; Score 70; DB 1; Length 303;
Best Local Similarity 35.2%; Pred. No. 1.7;
Matches 19; Conservative 10; Mismatches 17; Indels 8; Gaps 1;

QY / LEDWVINSDISISAFKALDUEZ IUSDAHAGAKAQTJLNS 32
| :: | 1::| |||| ::|| |||| :: | :| :| :|
Db 37 LNEUELLSSDISISVWTHIEEFKNVKFDKTIDSDTVKEIAIKLEQQLSKS 90

DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DNA REPAIR PROTEIN RADC HOMOLOG.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 15:13:38 ; Search time 10.24 seconds

(without alignments)
247.549 Million cell updates/sec

Title: US-09-367-009-3

Perfect score: 392

Sequence: 1 DSGCKLLEDVKTINSDIS.....TLKNFGLMHTVDSIWCNL 74

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	390	99.5	95	1	MGBB_HUMAN
2	208	53.1	93	1	MGBA_HUMAN
3	163	41.6	95	1	PSC3_RAT
4	70	17.9	303	1	FYSY_RICPR
5	66.5	17.0	234	1	RADC_HAEIN
6	65.5	16.7	131	1	YR59_MYCTU
7	65.5	16.7	603	1	VEI_HPV08
8	64.5	16.5	603	1	VEI_HPV21
9	64.5	16.5	611	1	DYNL_ARATH
10	61.5	15.7	162	1	IL15_MOUSE
11	61.5	15.7	570	1	HEML_KLJLA
12	61.5	15.7	734	1	MEPE_THEMA
13	61	15.6	605	1	VEI_HPV14
14	61	15.6	663	1	TA21_SCHPO
15	61	15.6	732	1	K086_MOUSE
16	60.5	15.4	430	1	SEPC_ARATH
17	60.5	15.4	684	1	TC10_YEAST
18	60	15.3	268	1	TRPC_ACICA
19	60	15.3	606	1	VEI_HPV5B
20	60	15.3	790	1	V90K_AMVLE
21	60	15.3	1071	1	PR16_YEAST
22	59.5	15.2	162	1	IL15_CERAE
23	59.5	15.2	162	1	IL15_HUMAN
24	59.5	15.2	162	1	IL15_MACMU
25	59.5	15.2	925	1	YB9G_SCHPO
26	59.5	15.2	1177	1	X307_MYCGE
27	59	15.1	604	1	VEI_HPV12
28	59	15.1	606	1	VEI_HPV05
29	59	15.1	690	1	Y023_NPVAC
30	59	15.1	2144	1	GL71_YEAST
31	59	15.1	4540	1	DYHC_PARTE
32	58.5	14.9	322	1	DIDH_RAT
33	58.5	14.9	606	1	RA17_SCHPO

ALIGNMENTS

RESULT 1

MGBB_HUMAN STANDARD; PRT; 95 AA.

AC 075556;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE MAMMAGLOBIN B PRECURSOR (MAMMAGLOBIN 2) (LIPOPHILIN C) (LACRYGLOBIN).

GN MGB2 OR UGB3 OR LIPHC.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE-99026127; PubMed-9806831;

RA Becker R.M., Darrow C., Zimonjic D.B., Popescu N.C., Watson M.A.,

RA Fleming T.P.;

RT "Identification of mammaglobin B, a novel member of the uteroglobin

RT gene family.";

RL Genomics 54:70-78(1998).

RN [2]

RP SEQUENCE FROM N.A.

RA MEDLINE-99167354; PubMed-10066439;

RA Zhao C., Nguyen T., Yusifov T., Glasgow B.J., Lehrer R.I.;

RT "Lipophilins: human peptides homologous to rat prostatein.";

RL Biochem. Biophys. Res. Commun. 256:147-155(1999).

RN [3]

RP SEQUENCE OF 19-85.

RT TISSUE-Tears;

RA MEDLINE-98163342; PubMed-9504814;

RA Molloy M.P., Bolis S., Herbert B.R., Ou K., Tyler M.I., van Dyk D.D.,

RA Wilcox M.D., Gooley A.A., Williams K.L., Morris C.A., Walsh B.J.;

RT "Establishment of the human reflex tear two-dimensional polyacrylamide

RT gel electrophoresis reference map: new proteins of potential

RT diagnostic value.";

RL Electrophoresis 18:2811-2815(1997).

RN [4]

RP SEQUENCE OF 19-46 AND 60-78, AND MASS SPECTROMETRY.

RA MEDLINE-98385871; PubMed-9720917;

RA Lehrer R.I., Xu G., Abduragimov A., Dinh N.N., Ou X.-D., Martin D.,

RA Glasgow B.J.;

RT "Lipophilin, a novel heterodimeric protein of human tears.";

RL FEBS Lett. 432:163-167(1998).

CC -!- FUNCTION: MAY BIND ANDROGENS AND OTHER STEROIDS, MAY ALSO BIND

CC ESTRADIOL, A CHEMOTHERAPEUTIC AGENT USED FOR PROSTATE CANCER.

CC MAY BE UNDER TRANSCRIPTIONAL REGULATION OF STEROID HORMONES.

CC -!- SUBUNIT: HETERODIMER OF A LIPOPHILIN A AND A LIPOPHILIN C

CC (MAMMAGLOBIN B) MONOMER ASSOCIATED HEAD TO HEAD.

CC -!- TISSUE SPECIFICITY: EXPRESSED IN THYMUS, TRACHEA, KIDNEY, STEROID

CC RESPONSIVE TISSUES (PROSTATE, TESTIS, UTERUS, BREAST AND OVARY),

CC AND SALIVARY GLAND.

CC -!- MASS SPECTROMETRY: MW=8854.94; METHOD=ELECTROSPRAY; RANGE=19-95.

CC -!- SIMILARITY: BELONGS TO THE UTEROGLOBIN FAMILY. LIPOPHILIN

CC SUBFAMILY.

CC

CC

CC

CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: AF071219; AAC79996.1; -
EMBL: AJ224173; CAA11865.1; -
MIM: 604396; -
InterPro: IPR000329; -
Pfam: PF01099; Uteroglobin; 1.
PROSITE: PS00403; UTEROGLBIN_1; FALSE_NEG.
PROSITE: PS00404; UTEROGLBIN_2; FALSE_NEG.
Signal: Glycoprotein.
SIGNAL 1 18
CHAIN 19 95 MAMMAGLOBIN B.
CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
SEQUENCE 95 AA: 10884 MW: 0719738289F8F8D CRC64;

Query Match 99.5%; Score 390; DB 1; Length 95;
Best Local Similarity 98.6%; Pred. No. 1.4e-35;
Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DSGCKLLEDVKTINSDISIPYKELLOEFIDSDAAEAMGKFCFLNQSHRTLNKFG 60
|||||
19 DSGCKLLEDVKTINSDISIPYKELLOEFIDSDAAEAMGKFCFLNQSHRTLNKFG 78

61 LMHTVYDSIWCNL 74
|||||
79 LMHTVYDSIWCNM 92

RESULT 2
ID MGBA_HUMAN STANDARD; PRT; 93 AA.
Q13296;
01-NOV-1997 (Rel. 35, Created)
01-NOV-1997 (Rel. 35, Last sequence update)
01-OCT-2000 (Rel. 40, Last annotation update)
MAMMAGLOBIN A PRECURSOR (MAMMAGLOBIN 1).
MGBl OR UGB2.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
TISSUE=Breast;
MEDLINE=96223698; PubMed=8631025;
Watson M.A., Fleming T.P.;
"Mammaglobin, a mammary-specific member of the uteroglobin gene family, is overexpressed in human breast cancer.";
Cancer Res. 56:860-865(1996).
[2]

SEQUENCE FROM N.A.
MEDLINE=98147371; PubMed=9488047;
Watson M.A., Darrow C., Zimonjic D.B., Popescu N.C., Fleming T.;
"Structure and transcriptional regulation of the human mamaglobin gene, a breast cancer associated member of the uteroglobin gene family localized to chromosome 11q13.";
Oncogene 16:817-824(1998).

-!- TISSUE SPECIFICITY: MAMMARY-SPECIFIC. OVER-EXPRESSED IN BREAST CANCER.

-!- SIMILARITY: BELONGS TO THE UTEROGLBIN FAMILY. LIPOPHILIN SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: U33147; AAC50375.1; -
EMBL: AF015224; AAC39608.1; -
MIM: 605562; -

InterPro: IPR000329; -
Pfam: PF01099; Uteroglobin; 1.
PROSITE: PS00403; UTEROGLBIN_1; FALSE_NEG.
PROSITE: PS00404; UTEROGLBIN_2; 1.

Signal: Glycoprotein.
SIGNAL 1 18 POTENTIAL.
CHAIN 19 93 MAMMAGLOBIN A.
CARBOHYD 53 53 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 93 AA: 10499 MW: 2895E8C43BF053E2 CRC64;

Query Match 53.1%; Score 208; DB 1; Length 93;
Best Local Similarity 53.4%; Pred. No. 5.9e-16;
Matches 39; Conservative 12; Mismatches 22; Indels 0; Gaps 0;

QY 2 SGCKLLEDVKTINSDISIPYKELLOEFIDSDAAEAMGKFCFLNQSHRTLNKFG 61
|||||
20 SGCKLLEDVKTINSDISIPYKELLOEFIDSDAAEAMGKFCFLNQSHRTLNKFG 79

QY 62 MMHTVYDSIWCNL 74
|||||
80 FMQLIYDSSLCDL 92

RESULT 3
ID PSC3_RAT STANDARD; PRT; 95 AA.
AC P02780; Q63463;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE PROSTATIC STEROID-BINDING PROTEIN C3 CHAIN PRECURSOR (PROSTATEIN PEPTIDE C3).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
NCBI_TaxID=10116;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=83082848; PubMed=6294095;
Parker M.G., White R., Hurst H., Needham M., Tilly R.;
"Prostatic steroid-binding protein. Isolation and characterization of C3 genes.";
J. Biol. Chem. 258:12-15(1983).
[2]
SEQUENCE FROM N.A.
RX MEDLINE=83238526; PubMed=6190812;
Viskochil D.H., Perry S.T., Lea O.A., Stafford D.W., Wilson E.M., French F.S.;
"Isolation of two genomic sequences encoding the Mr = 14,000 subunit of rat prostatein.";
J. Biol. Chem. 258:8861-8866(1983).
[3]
SEQUENCE FROM N.A.
RX MEDLINE=92165796; PubMed=1537831;
Tan J.A., Marschke K.B., Ho K.-C., Perry S.T., Wilson E.M., French F.S.;
"Response elements of the androgen-regulated C3 gene.";
J. Biol. Chem. 267:4456-4466(1992).
[4]
RP ERRATUM.
RX MEDLINE=92218467; PubMed=1339454;
Tan J.A., Marschke K.B., Ho K.C., Perry S.T., Wilson E.M., French F.S.;
J. Biol. Chem. 267:7958-7958(1992).

```

[5]
RN  SEQUENCE OF 19-95.
RP  MEDLINE=8118769; PubMed=7014218;
RA  Pesters B., Rombauts W., Mous J., Heyns W.;
RT  "Structural studies on rat prostatic binding protein. The primary
RT  structure of its glycosylated component C3.";
RL  Eur. J. Biochem. 115:115-121(1981).
CC  -!- FUNCTION: PART OF PROSTATEIN WHICH IS THE MAJOR SECRETORY
CC  GLYCOPROTEIN OF VENTRAL PROSTATE GLAND. STEROID-BINDING PROTEIN;
CC  CAN BIND NON-POLAR STEROIDS, CHOLESTEROL AND A GROUP OF SMALL
CC  PROLINE-RICH PEPTIDES.
CC  -!- SUBUNIT: PROSTATEIN IS COMPOSED OF THREE DIFFERENT PEPTIDES CALLED
CC  C1, C2 AND C3. THESE FORM COVALENT C1:C3 (F) AND C2:C3 (S)
CC  HETERODIMERS WHOSE NONCOVALENT ASSOCIATION FORMS TETRAMERIC
CC  (C1:C3/C3:C2) PROSTATEIN MOLECULES.
CC  -!- SUBCELLULAR LOCATION: SECRETED.
CC  -!- TISSUE SPECIFICITY: VENTRAL PROSTATE.
CC  -!- INDUCTION: ANDROGEN DEPENDENT, AS SHOWN BY THE DECREASE IN THE
CC  LEVEL OF THE PROTEIN FOLLOWING CASTRATION.
CC  -!- MISCELLANEOUS: C3 IS ENCODED BY TWO DIFFERENT GENES.
CC  -!- SIMILARITY: BELONGS TO THE UTEROGLOBIN FAMILY. LIPOPHILIN
CC  SUBFAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
EMBL; V01263; CAA24577.1; -
EMBL; M71245; AAA41965.1; -
PIR; A03250; B0RT3.
InterPro: IPR000329; -
Pfam: PF01099; Uterogloblin; 1.
PROSITE; PS00403; UTEROGLOBIN_1; FALSE_NEG.
PROSITE; PS00404; UTEROGLOBIN_2; 1.
Signal; Glycoprotein; Steroid-binding.
SIGNAL 1 18
CHAIN 19 95
PROSTATIC STEROID-BINDING PROTEIN C3
CHAIN.
CARBOHYD 35 35
CONFLICT 53 53
CONFLICT 79 79
SEQUENCE 95 AA; 10730 MW; F7F7F1A0C982E375 CRC64;

Query Match 41.6%; Score 163; DB 1; Length 95;
Best Local Similarity 36.6%; Pred. No. 4.3e-11;
Matches 26; Conservative 25; Mismatches 20; Indels 0; Gaps 0;

2 SGCKLLEDVKEKTSNDISIPYKELQFIDSDAAAEAMGKFCFLNQSHRTLNKFL 61
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
21 SGCSILDEVIRGTINSTVTLHDYMLKVPQVDHTEKAVKQKFCFLQDTKLENVGV 80
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

62 MHTVVDYSWC 72
|||:|||||:
81 MMEAFNSESC 91

Query Match 17.9%; Score 70; DB 1; Length 303;
Best Local Similarity 35.2%; Pred. No. 1.7;
Matches 19; Conservative 10; Mismatches 17; Indels 8; Gaps 1;

QY 7 LEDMWKTSNDISIPYKELQF-----IDSDAAAEAMGKFCFLNQS 52
DB 37 LNELELLISDISISVWTHIIIEEFKNVFKDTIDSVTVEAKLIEQOLSKS 90
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 5
ID RADC_HAEIN STANDARD; PRT; 234 AA.
AC P44952;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DNA REPAIR PROTEIN RADC HOMOLOG.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
```

```

NCBI_TaxID=782;
[1]
RN  SEQUENCE FROM N.A.
RP  STRAIN=MADRID E;
RX  MEDLINE=97419517; PubMed=9274032;
RA  Andersson J.O., Andersson S.G.E.;
RT  "Genomic rearrangements during evolution of the obligate
RT  intracellular parasite Rickettsia prowazekii as inferred from an
RT  analysis of 52015 bp nucleotide sequence.";
RL  Microbiology 143:2783-2795(1997).
[2]
RN  SEQUENCE FROM N.A.
RP  STRAIN=MADRID E;
RX  MEDLINE=99039499; PubMed=9823893;
RA  Andersson S.G.E., Zomrodipour A., Andersson J.O.,
RA  Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA  Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT  "The genome sequence of Rickettsia prowazekii and the origin of
RT  mitochondria";
RL  Nature 396:133-140(1998).
[3]
RN  SEQUENCE OF 168-303 FROM N.A.
RP  STRAIN=MADRID E, AND B;
RX  MEDLINE=99416441; PubMed=10486973;
RA  Andersson J.O., Andersson S.G.E.;
RT  "Genome degradation is an ongoing process in Rickettsia.";
RL  Mol. Biol. Evol. 16:1178-1191(1999).
CC  -!- FUNCTION: FUNCTIONAL HOMOLOG OF SRP RECEPTOR. PROBABLY INVOLVED IN
CC  THE RECEPTION AND INSERTION OF A SUBSET OF PROTEINS AT THE
CC  MEMBRANE (BY SIMILARITY).
CC  -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED (BY SIMILARITY).
CC  -!- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
EMBL; Y11784; CAA72477.1; -
EMBL; AJ235273; CAA15202.1; -
EMBL; AJ238755; CAB56084.1; -
EMBL; AJ238756; CAB56088.1; -
InterPro: IPR000897; -
Pfam: PF00448; SRP54; 1.
PROSITE; PS00300; SRP54; 1.
Signal recognition particle; GTP-binding; RNA-binding; Membrane;
Cell division.
NP_BIND 108 115
NP_BIND 190 194
NP_BIND 254 257
SEQUENCE 303 AA; 33313 MW; BF3F7EB9383B7007 CRC64;

Query Match 17.9%; Score 70; DB 1; Length 303;
Best Local Similarity 35.2%; Pred. No. 1.7;
Matches 19; Conservative 10; Mismatches 17; Indels 8; Gaps 1;

QY 7 LEDMWKTSNDISIPYKELQF-----IDSDAAAEAMGKFCFLNQS 52
DB 37 LNELELLISDISISVWTHIIIEEFKNVFKDTIDSVTVEAKLIEQOLSKS 90
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 5
ID RADC_HAEIN STANDARD; PRT; 234 AA.
AC P44952;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DNA REPAIR PROTEIN RADC HOMOLOG.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
```



```

DR Pfam; PF00519; El; 1.
DR Pfam; PF00524; El_N; 1.
KW Early protein; DNA replication; Helicase; ATP-binding; DNA-binding;
KW Nuclear protein.
FT NP_BIND 431 438 ATP (POTENTIAL).
SQ SEQUENCE 603 AA; 68821 MW; 0813860098D48AD CRC64;

Query Match
Best Local Similarity 16.7%; Score 65.5; DB 1; Length 603;
Matches 22; Conservative 21; Mismatches 23; Indels 33; Gaps 4;

Oy 1 DSGCKLLEDWVEKINSISIP-----EYKELLOEFIDSDAAEAMGK 43
Dg 112 DSGVELTLNNEAEVDSVEVPAIDSRPDEGGGALDIDYATLLR---SSNTRATLMK 168
Dy 44 FKQCFLN-----QSHRTLNKFLM-----MHTVYDS 69
Dg 169 FKEAFGDFGNELTROFKSYKTCNYYVVAAYAVHDVYES 207

RESULT 8
REL_HPV21
ID VEL_HPV21 STANDARD; PRT; 603 AA.
AC P50759;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE REPLICATION PROTEIN EI.
OS El.
OC Human papillomavirus type 21.
OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
NCBI_TaxID=31548;
CC SEQUENCE FROM N.A.
CC Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ATP-DEPENDENT DNA HELICASE REQUIRED FOR INITIATION OF
CC VIRAL DNA REPLICATION. IT FORMS A COMPLEX WITH THE VIRAL E2
CC PROTEIN. THE E1-E2 COMPLEX BINDS TO THE REPLICATION ORIGIN WHICH
CC CONTAINS BINDING SITES FOR BOTH PROTEINS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L31779; AAA79396.1;
CC InterPro: IPR001177;
CC Pfam; PF00519; El; 1.
CC Pfam; PF00524; El_N; 1.
CC Early protein; DNA replication; Helicase; ATP-binding; DNA-binding;
CC Nuclear protein.
CC NP_BIND 431 438 ATP (POTENTIAL).
CC SEQUENCE 603 AA; 68983 MW; 93006494BEF10339 CRC64;

Query Match
Best Local Similarity 16.5%; Score 64.5; DB 1; Length 603;
Matches 26; Conservative 15; Mismatches 23; Indels 35; Gaps 6;

Oy 1 DSG--CKLLEDWVEKINSISIP-----EYKELLOEFIDSDAAEAM 41
Dg 112 DSGLECLTN---EDVSEVEVPALDSQPVAAEQLTGVDIHYKELLRA---SNNKATILM 165
Oy 42 GKFKOCF---LNQSHRTLNKFG-----LMHTVYDSI 70
Dg 166 AKFEFFGVGNDLTROFKSYKTCNANVLSVYAVHDDL 204

```

Copied from PCF009367009 on 27/04/2004

```

RESULT 9
DYNL_ARATH
ID DYNL_ARATH STANDARD; PRT; 611 AA.
AC P42697;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DYNAMIN-LIKE PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; Rosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OC NCBI_TaxID=3702;
CC (1)
CC SEQUENCE FROM N.A.
CC TISSUE=Leaf;
CC Hwang I., Goodman H.M., Yoon H.W., Yoon J.-H., Cho M.J.;
CC "Cloning and characterization of a cDNA clone encoding dynamin-like
CC GTP binding protein in Arabidopsis thaliana.";
CC Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE DYNAMIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L36939; AAB63528.1;
CC InterPro: IPR000375;
CC InterPro: IPR001401;
CC Pfam; PF00350; dynamin; 1.
CC Pfam; PF01031; dynamin_2; 1.
CC PRINTS; PR00195; DYNAMIN.
CC PROSITE; PS00410; DYNAMIN; 1.
CC Motor protein; GTP-binding.
CC NP_BIND 41 48 GTP (POTENTIAL).
CC NP_BIND 142 146 GTP (POTENTIAL).
CC NP_BIND 211 214 GTP (POTENTIAL).
CC SEQUENCE 611 AA; 68509 MW; EDE98B79B1BC0D5 CRC64;

Query Match
Best Local Similarity 16.5%; Score 64.5; DB 1; Length 611;
Matches 17; Conservative 21; Mismatches 23; Indels 21; Gaps 3;

Oy 6 LLEDVKTINSISIPYKELLOEFIDSDAAEAMGKFK-----QC----- 47
Dg 432 ILKDLVHKSVNETVKQYPALRVEV--TNAIESLDKMGSKKATQLQVDMCSYLTV 489
Oy 48 -FLNQSHRTLNKFLMHTVYD 68
Dg 490 DFFRKLQPDQVEKGNPHTSIFD 511

RESULT 10
IL15_MOUSE
ID IL15_MOUSE STANDARD; PRT; 162 AA.
AC P48346;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE INTERLEUKIN-15 PRECURSOR (IL-15).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
CC (1)
CC SEQUENCE FROM N.A.

```


Search completed: April 24, 2001, 15:14:21
Job time: 43 sec

Matches 18: Conservative 19: Mismatches 29: Indels 12: Gaps 4:

Oy 4 CKLLDMYKTKINSISIPYKELQEFIDSDAAAEAMGK--FKOCFLNQSHRTLKNEFL 61
Db 141 CONIANMAEQKVMVSAI--FSESSKDIVNPESFSEGLGKTVKDLKDYEFNEQLTK-YGL 197
Oy 62 MMHTV-----YDSIWC 72
Db 198 EPTIFFSVIRKYDAYWC 215

RESULT 15
KUB6_MOUSE STANDARD; PRT; 732 AA.
P27641;
01-AUG-1992 (Rel. 23, Created)
01-AUG-1992 (Rel. 23, Last sequence update)
30-MAY-2000 (Rel. 39, Last annotation update)
ATP-DEPENDENT DNA HELICASE II, 80 KDA SUBUNIT (KU AUTOANTIGEN PROTEIN
P86 HOMOLOG) (KUB80) (CTC BOX BINDING FACTOR 85 KDA SUBUNIT) (CTCBF)
(CTC85) (NUCLEAR FACTOR IV) (DNA-REPAIR PROTEIN XRCC5).
XRCC5 OR G22P2.
Mus musculus (Mouse).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]

SEQUENCE FROM N.A.
STRAIN=BALB/C;
MEDLINE=92350687; PubMed=1641347;
Falzon M., Kuff E.L.;
"The nucleotide sequence of a mouse cDNA encoding the 80 kDa subunit
of the Ku (p70/p80) autoantigen.";
Nucleic Acids Res. 20:3784-3784(1992).
-!- FUNCTION: SINGLE STRANDED DNA-DEPENDENT ATP-DEPENDENT HELICASE.
HAS A ROLE IN CHROMOSOME TRANSLLOCATION. THE DNA HELICASE II
COMPLEX BINDS PREFERENTIALLY TO FORK-LIKE ENDS OF DOUBLE-STRANDED
DNA IN A CELL CYCLE-DEPENDENT MANNER. IT WORKS IN THE 3'-5'
DIRECTION. BINDING TO DNA MAY BE MEDIATED BY P70.
-!- SUBUNIT: HETERODIMER OF A 70 KDA AND A 80 KDA SUBUNIT.
-!- SUBCELLULAR LOCATION: NUCLEAR.
-!- PTM: PHOSPHORYLATED IN VIVO AT SERINE RESIDUES (BY SIMILARITY).

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; X66323; CAA46999.1; -.
PIR; S24943; S24943.
PIR; S26303; S26303.
DR MGD; MGI:104517; xrcc5.
Nuclear protein; DNA-binding; Phosphorylation; Helicase.
DOMAIN 138 165 LEUCINE-ZIPPER.
FT DOMAIN 478 520 PRO-RICH.
FT MOD_RES 651 651 PHOSPHORYLATION (BY NUCLEAR KINASE NII)
(POTENTIAL).
SEQUENCE 732 AA; 83323 MW; 06A2270A0ECDDCC3 CRC64;

Query Match 15.6%; Score 61; DB 1; Length 732;
Best Local Similarity 32.7%; Pred. No. 42;
Matches 17; Conservative 6; Mismatches 17; Indels 12; Gaps 1;

Oy 16 NSDISIPEYKELQEFIDSDAAAEAMGKFKOCFLNQSHRTLKNEFLMMHTVY 67
Db 256 NLSIKVAYKSIVQE-----KFKKSMVVVDARTLAKEDIQETVY 295

OM of: US-09-367-009-3 to: Issued_Patents_NA.* out_format : pfs
 Date: Apr 28, 2001 4:46 PM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL-frame+ p2n.model -DRV=xlp
 -Q/cgn2_1/USPTO.spool/US09367009/runat_24042001_151807_19038/app_query.fasta_1.131
 -DB-Issued_Patents_NA -OFMT=fastap -SUFFIX=rni -GAPOP=12.000
 -GAEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -OGAPOP=4.500 -OGAEXT=0.050 -XGAPOP=10.000 -XGAEXT=0.500
 -FGAPOP=6.000 -FGAEXT=7.000 -YGAPOP=10.000 -YGAEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
 -TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -NORM-ext -MINLEN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -USER=US09367009 @CGN1_152 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:

Query: US-09-367-009-3
 Query length: 74
 Database: Issued_Patents_NA.*
 Database sequences: 302621
 Search time (sec): 80.170000

core_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-821-451A-5 +	390.00	1031.08	7.6e-50	476	1
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-263-810-5 +	390.00	1031.08	7.6e-50	476	1
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-455-896-5 +	208.00	543.02	1.2e-22	403	1
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-933-149-5 +	208.00	543.02	1.2e-22	403	1
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-09-082-343-5 +	208.00	543.02	1.2e-22	403	1
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-09-082-253-5 +	208.00	543.02	1.2e-22	403	1
/cgn2_6/ptodata/2/ina/PCPUS.COMB.seq:US-08-947-823-4 +	208.00	543.02	1.2e-22	403	1
/cgn2_6/ptodata/2/ina/PCPUS.COMB.seq:US-08-947-823-4 +	208.00	543.02	1.2e-22	403	1
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-455-896-1 +	208.00	540.67	1.6e-22	503	1
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-933-149-1 +	208.00	540.67	1.6e-22	503	1
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-09-082-343-1 +	208.00	540.67	1.6e-22	503	1
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-09-082-253-1 +	208.00	540.67	1.6e-22	503	1
/cgn2_6/ptodata/2/ina/PCPUS.COMB.seq:US-08-947-823-4 +	208.00	540.67	1.6e-22	503	1
/cgn2_6/ptodata/2/ina/PCPUS.COMB.seq:US-08-947-823-4 +	208.00	540.67	1.6e-22	503	1
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-08-947-823-1 +	61.50	97.18	792.11	51952	1
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-933-149-1 +	61.00	127.97	15.27	2515	1
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-933-149-1 +	60.50	130.78	10.65	1700	1
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-252-329-4 +	60.50	130.78	10.65	1700	1
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-393-305-12 +	59.50	145.01	1.72	345	1
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-393-305-13 +	59.50	145.01	1.72	345	1
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-393-305-12 +	59.50	145.01	1.72	345	1
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-393-305-13 +	59.50	145.01	1.72	345	1
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-725-969-12 +	59.50	145.01	1.72	345	1
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-725-969-13 +	59.50	145.01	1.72	345	1
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-725-969-12 +	59.50	145.01	1.72	345	1
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-725-969-13 +	59.50	145.01	1.72	345	1
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-794-524-12 +	59.50	145.01	1.72	345	1
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-794-524-12 +	59.50	145.01	1.72	345	1
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-189-193-12 +	59.50	145.01	1.72	345	1
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-189-193-12 +	59.50	145.01	1.72	345	1
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-031-399-1 +	59.50	141.31	2.76	489	1
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-031-399-1 +	59.50	141.31	2.76	489	1
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-393-305-1 +	59.50	141.31	2.76	489	1
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-393-305-4 +	59.50	141.31	2.76	489	1
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-535-733-1 +	59.50	141.31	2.76	489	1
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-726-817-1 +	59.50	141.31	2.76	489	1
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-726-817-1 +	59.50	141.31	2.76	489	1
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-504-042-1 +	59.50	141.31	2.76	489	1
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-504-042-1 +	59.50	141.31	2.76	489	1
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-392-317B-1 +	59.50	141.31	2.76	489	1
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-392-317B-2 +	59.50	141.31	2.76	489	1
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-725-969-1 +	59.50	141.31	2.76	489	1
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-725-969-4 +	59.50	141.31	2.76	489	1

/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-794-524-1 + 59.50 141.31 2.76
 /cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-794-524-4 + 59.50 141.31 2.76
 /cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-842-947-5 + 59.50 141.31 2.76
 /cgn2_6/ptodata/2/ina/6A.COMB.seq:US-09-134-132-1 + 59.50 141.31 2.76
 /cgn2_6/ptodata/2/ina/6A.COMB.seq:US-09-134-132-2 + 59.50 141.31 2.76
 seq_name: /cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-821-451A-5
 seq_documentation_block:
 ; Sequence 5, Application US/08821451A
 ; Patent No. 6066724
 ; GENERAL INFORMATION:
 ; APPLICANT: Jian Ni, Guo-Liang Yu and Reiner Gentz
 ; TITLE OF INVENTION: Human Endometrial Specific Steroid-
 ; NUMBER OF SEQUENCES: 27
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
 ; ADDRESSEE: CECCHI, STEWART & OLSTEIN
 ; STREET: 6 BECKER FARM ROAD
 ; CITY: ROSELAND
 ; STATE: NEW JERSEY
 ; COUNTRY: USA
 ; ZIP: 07068
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5 INCH DISKETTE
 ; COMPUTER: IBM PS/2
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: WORD PERFECT 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/821.451A
 ; FILING DATE: March 21, 1997
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION NUMBER: 60/014.724
 ; FILING DATE: March 21, 1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: MULLINS, J.G.
 ; REGISTRATION NUMBER: 33,073
 ; REFERENCE/DOCKET NUMBER: 325800-521 (PF257)
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 201-994-1700
 ; TELEFAX: 201-994-1744
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 476 BASE PAIRS
 ; TYPE: NUCLEIC ACID
 ; STRANDEDNESS: SINGLE
 ; TOPOLOGY: LINEAR
 ; MOLECULE TYPE: CDNA
 ; US-08-821-451A-5
 alignment_scores:
 Quality: 390.00 Length: 74
 Ratio: 5.270 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 98.649
 alignment_block:
 US-09-367-009-3 x US-08-821-451A-5
 Align seg 1/1 to: US-08-821-451A-5 from: 1 to: 476
 1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
 100 GATTCTGGTGGCAACTCTCGGAGGACATGGTTGAAAGACCATCAATTC 149
 17 rAspIleSerIleProGluTyrlYsGluLeuGluGlnGluPheIleAspS 34
 150 CGACATATCTATACCTGGAATCAAGAGCTTCTCAAGAGTTTCATAGACA 199
 34 erAspIleAlaIalaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
 |||||||||||||||||||||||||||||||||||||||||||||||||||

200 GTGATGCCGTCAGAGGCTATGGGAAATTCAGAGAGTGTTCCTCAAC 249
51 GlnSerHisArgThrLeuLysAsnGlyLeuMetMetHisThrValty 67
|||||
250 CAGTCACATAGAACTCTGAAAACCTTTGGACTGATGATGCATACAGTGA 299
67 rAspSerIleTrpCysAsnLeu 74
|||||
300 CGACAGCATTTGGTGAATATG 321

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-09-263-810-5

seq_documentation_block:
; Sequence 5, Application US/09263810
; Patent No. 6174992
GENERAL INFORMATION:
APPLICANT: JIAN NI, GUO-LIANG YU AND REINER GENTZ
TITLE OF INVENTION: Human Endometrial Specific Steroid-
Binding Factor I, II and III
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELIA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/263,810
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/821,451
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: MULLINS, J.G.
REGISTRATION NUMBER: 33,073
REFERENCE/DOCKET NUMBER: 325800-521 (PF257)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 476 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
US-09-263-810-5

alignment_scores:
Quality: 390.00 Length: 74
Ratio: 5.270 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.649

alignment_block:

US-09-367-009-3 x US-09-263-810-5

Align seg 1/1 to: US-09-263-810-5 from: 1 to: 476

1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
|||||
100 GATTCTCGCTGCAAACTCTCGGAGGACATGTTGAAAGACCATCAATC 149
17 rAspIleSerIleProGluTyrLysGluLeuGluGluPheIleAsps 34
|||||

150 CGACATATATCTACTGTAATACAAAGAGCTTCTTCAAGAGTTCATAGACA 199
34 exAspAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
|||||
200 GTGATGCCGTCAGAGGCTATGGGAAATTCAGAGAGTGTTCCTCAAC 249
51 GlnSerHisArgThrLeuLysAsnGlyLeuMetMetHisThrValty 67
|||||
250 CAGTCACATAGAACTCTGAAAACCTTTGGACTGATGATGCATACAGTGA 299
67 rAspSerIleTrpCysAsnLeu 74
|||||
300 CGACAGCATTTGGTGAATATG 321

seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-08-455-896-5

seq_documentation_block:
; Sequence 5, Application US/08455896
; Patent No. 5668267
GENERAL INFORMATION:
APPLICANT: WATSON, MARK A.
APPLICANT: FLEMING, TIMOTHY P.
TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: USA
ZIP: 63103-1817
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455,896
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: HOLLAND, DONALD R.
REGISTRATION NUMBER: 35,197
REFERENCE/DOCKET NUMBER: 952726
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 727-5188
TELEFAX: (314) 727-6092
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 403 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-455-896-5

alignment_scores:
Quality: 208.00 Length: 73
Ratio: 3.355 Gaps: 0
Percent Similarity: 84.932 Percent Identity: 53.425

alignment_block:

US-09-367-009-3 x US-08-455-896-5

Align seg 1/1 to: US-08-455-896-5 from: 1 to: 403

2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
|||||
118 TCTGGCTGCCCTTATTGGAGATGTTTCCAGACAATCAATCCACA 167

ignment_block:
US-09-367-009-3 x US-08-933-149-5 ..
align seq. 1/1 to: US-08-933-149-5 from: 1 to: 403

;

```

alignment_block:
US-09-367-009-3 x US-09-082-253-5
Align seg 1/1 to: US-09-082-253-5 from: 1 to: 403
2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
118 TCCTGGCTGCCCTTATTGGAGAAATGATTCACAGACAATCAATCCACA 167
18 pIleSerIleProGluTyrLysGluLeuGluPheIleAspSerA 35
168 AGTGCTTAAGACTGAATACAAAGAACTTCTCAAGAGATTCATAGACGACA 217
35 spAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
218 ATGCCACTACAATGCCATAGATGAATGAAGGAATGTTTCTTAACCAA 267
52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
268 ACGGATGAACCTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATATGA 317
68 pSerIleTrpCysAsnLeu 74
318 CAGCAGCTCTTTGTGATTGA 336
seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq: PCT-US96-08235-5
seq_documentation_block:
; Sequence 5, Application PC/TUS9608235
; GENERAL INFORMATION:
; APPLICANT: WATSON, MARK A.
; TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
; TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 63105-1817
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/08235
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/455,896
; FILING DATE: 05/31/1995
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 964796
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ANTI-SENSE: NO
; HYPOTHETICAL: NO

```

```

alignment_block:
US-09-367-009-3 x US-09-082-343-5
Align seg 1/1 to: US-09-082-343-5 from: 1 to: 403
2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
118 TCCTGGCTGCCCTTATTGGAGAAATGATTCACAGACAATCAATCCACA 167
18 pIleSerIleProGluTyrLysGluLeuGluPheIleAspSerA 35
168 AGTGCTTAAGACTGAATACAAAGAACTTCTCAAGAGATTCATAGACGACA 217
35 spAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
218 ATGCCACTACAATGCCATAGATGAATGAAGGAATGTTTCTTAACCAA 267
52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
268 ACGGATGAACCTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATATGA 317
68 pSerIleTrpCysAsnLeu 74
318 CAGCAGCTCTTTGTGATTGA 336
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq: US-09-082-253-5
seq_documentation_block:
; Sequence 5, Application US/09082253
; Patent No. 6004756
; GENERAL INFORMATION:
; APPLICANT: WATSON, MARK A.
; APPLICANT: FLEMING, TIMOTHY P.
; TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
; TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 63105-1817
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/082,253
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/455,896
; FILING DATE: 05/31/1995
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 952726
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ANTI-SENSE: NO
; HYPOTHETICAL: NO

```

```

; ANTI-SENSE: NO
PCT-US96-08235-5

alignment_scores:
  Quality: 208.00      Length: 73
  Ratio: 3.355        Gaps: 0
  Percent Similarity: 84.932  Percent Identity: 53.425

alignment_block:
US-09-367-009-3 x PCT-US96-08235-5 ..
  Align seg 1/1 to: PCT-US96-08235-5 from: 1 to: 403
2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
118 TCTGGCTGCCCTTATTGGAGATGTGATTTCCAAGACAATCAATCCACA 167
18 pIleSerIleProGluTyrLysGluLeuGlnGluPheIleAspSerA 35
168 AGTGTCTAAGACTGAATACAAAGAACTTCTCAAGAGTTTCATAGACGACA 217
35 spLaLaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
218 ATGCCACTACAAATGCCATAGATGATTTGAAGGAATGTTTCTTAACCAA 267
52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
268 ACGATGAACACTCTGACCAATGTGAGGTGTTTATGCAATTAATATGA 317
68 pSerIleTyrCysAsnLeu 74
318 CAGCAGTCTTTGTGATTTA 336

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-455-896-1

seq_documentation_block:
; Sequence 1, Application US/08455896
; Patent No. 5668267
; GENERAL INFORMATION:
; APPLICANT: WATSON, MARK A.
; APPLICANT: FLEMING, TIMOTHY P.
; TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
; TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 63105-1817
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/455,896
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 952726
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 503 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-455-896-1

alignment_scores:
  Quality: 208.00      Length: 73
  Ratio: 3.355        Gaps: 0
  Percent Similarity: 84.932  Percent Identity: 53.425

alignment_block:
US-09-367-009-3 x US-08-455-896-1 ..
  Align seg 1/1 to: US-08-455-896-1 from: 1 to: 503
2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
118 TCTGGCTGCCCTTATTGGAGATGTGATTTCCAAGACAATCAATCCACA 167
18 pIleSerIleProGluTyrLysGluLeuGlnGluPheIleAspSerA 35
168 AGTGTCTAAGACTGAATACAAAGAACTTCTCAAGAGTTTCATAGACGACA 217
35 spLaLaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
218 ATGCCACTACAAATGCCATAGATGATTTGAAGGAATGTTTCTTAACCAA 267
52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
268 ACGATGAACACTCTGACCAATGTGAGGTGTTTATGCAATTAATATGA 317
68 pSerIleTyrCysAsnLeu 74
318 CAGCAGTCTTTGTGATTTA 336

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-933-149-1

seq_documentation_block:
; Sequence 1, Application US/08933149
; Patent No. 5922836
; GENERAL INFORMATION:
; APPLICANT: WATSON, MARK A.
; APPLICANT: FLEMING, TIMOTHY P.
; TITLE OF INVENTION: MAMMAGLOBIN, A SECRETED
; TITLE OF INVENTION: MAMMARY SPECIFIC BREAST CANCER PROTEIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 63105-1817
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,149
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: HENDERSON, MELODIE W.
; REGISTRATION NUMBER: 37,848
; REFERENCE/DOCKET NUMBER: 6029-6040
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

```



```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/08235
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 964796
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 503 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PCT-US96-08235-1

```

alignment_scores:			
Quality:	208.00	Length:	73
Ratio:	3.355	Gaps:	0
Percent Similarity:	84.932	Percent Identity:	53.425

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
||||| ||||| ::::: ||||| |||||
118 TCTGGCTGCCCCCTATTGGAGATGTGATTCCCAAGACAATCAATCCCACA 167

[illegible]

35 spAlaAalaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
:: ::|||::||| :::::||:: ::|||

68 pserIleTrpCysAsnLeu 74
||||| |||::|||

318 CAGAGCTTTGTGATTA 336

seq_name: /cqn2_6/ptodata/2/ina/PCTUS_COMB.seq: PCT-US96-08235-1

seq_documentation_block:

Sequence 1, Application PC/TUS9608235

GENERAL INFORMATION:

APPLICANT: WATSON, MARK A.
APPLICANT: ELEMING TIMOTHY B

APPLICANT: FLEMING, TIMOTHY P.
TITLE OF INVENTION: DNA SEQUENCE

TITLE OF INVENTION: MAMMARY-SPECIFIC

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESS: ROGERS, HOWELL & HAF
STREET. 7733 FORSYTH BOULEVARD

CITY: ST. LOUIS

STATE: MISSOURI

COUNTRY: USA

ZIP: 63105-1817
COMPUTER READABLE FORM.

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

Copied from PCT009367009 on 27-04-2004

COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/947,823
FILING DATE: 09-OCT-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/18802
FILING DATE: 09-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,191
FILING DATE: 10-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 023070-070210US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3982 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
FEATURE:
NAME/KEY: CDS
LOCATION: 87..3860
OTHER INFORMATION: /note= "Copy 2 cdna for M1 nematode
OTHER INFORMATION: resistance gene of tomato"
US-08-947-823-4

alignment_scores:	
Quality:	61.50
Ratio:	1.757
Percent Similarity:	67.308
Percent Identity:	28.846
Length:	52
Gaps:	1

```

alignment_block:
US-09-367-009-3 x US-08-947-823-4 ..
Align seg 1/1 to: US-08-947-823-4 from: 1 to: 3982
23 GluTyrLysGluLeuLeuGlnGluPheIleAspSerAspAlaAlaG1 39
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
321 GAGGTTGAGAATCTGCTTCAATCACTTTTGGATGATGATGCTCTTACTAG 370
..
39 uAlaMetGlyLysPheLysGlnCysPheLeuAsnGlnSerHisArgThrL 56
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
371 CCTCACCAGTAATATGGATGACTGT...ATCAGCTTGTATCATCGTCTTT 417
..
56 euLysAsnPheGlyLeuMetMetHisThrValTyrAspSerIleTrrPcys 72
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
418 ATAATCAGATGCCATCATGATGATGAGCAATTGGACTTCCTCCTCTTG 467
..
73 AsnLeu 74
|||||
468 AATCTG 473
..
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-947-823-1
seq_documentation_block:
; Sequence 1, Application US/08947823
; Patent No. 6114605
; GENERAL INFORMATION:
; APPLICANT: Williamson, Valerie M.
; APPLICANT: Kaloshian, Isquohi

```

```

seq_documentation_block:
; Sequence 1, Application US/08695866
; Patent No. 5955644
; GENERAL INFORMATION:
; APPLICANT: Hasty, Paul
; APPLICANT: Lim, Dae-Sik
; TITLE OF INVENTION: KU-DEFICIENT CELLS AND NON-HUMAN
; TITLE OF INVENTION: TRANSGENIC ANIMALS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of The Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/695,866
; FILING DATE: 08-AUG-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 008535-0011-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2515 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-695-866-1

alignment_scores:
Quality: 61.00 Length: 52
Ratio: 2.033 Gaps: 1
Percent Similarity: 57.692 Percent Identity: 32.692

alignment_block:
US-09-367-009-3 x US-08-695-866-1 ..

Align seg 1/1 to: US-08-695-866-1 from: 1 to: 2515

16 AsnSerAspIleSerIleProGluTyrLysGluLeuGlnGluPheIl 32
||| ::::::::::| | | | | | | | | | | | | | | | | | | |
793 AACTTGCTATAAGATGTAGCCTATAAATCGATTGTACAGGAG..... 837

32 eAspSerAspAlaAlaGluAlaMetGlyLysPheLysGlnCysPheL 49
| | | | | | | | | | | | | | | | | | | | | | | | | |
838 .....AATTTAAGAGAGTTGGG 856

49 euAsnGlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThr 65
: : : | | | | | | | | | | | | | | | | | | | | | |
857 TAGTTGTGGAGCAAGACCGCTAAAGAGGAGAGATATACAAAAGAGACT 906

66 ValTyr 67
| | | | |
907 GTTAT 912

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 15:12:13 ; Search time 12.98 Seconds
(without alignments)
109.523 Million cell updates/sec

Title: US-09-367-009-3

Perfect score: 392
Sequence: 1 DSGCKLLEDMVEKINSDIS.....TLKNFLMHTVYDSTWNL 74

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA.*

1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	390	99.5	95	3	US-08-821-451A-6
2	390	99.5	95	4	US-08-263-810-6
3	208	53.1	93	1	US-08-455-896-2
4	208	53.1	93	2	US-08-933-149-2
5	208	53.1	93	2	US-09-082-343-2
6	208	53.1	93	3	US-09-082-253-2
7	208	53.1	93	5	PCT-US96-08235-2
8	164	41.8	95	3	US-08-821-451A-27
9	164	41.8	95	4	US-08-263-810-27
10	163	41.6	95	1	US-08-455-896-7
11	163	41.6	95	2	US-08-933-149-7
12	163	41.6	95	2	US-09-082-343-7
13	163	41.6	95	3	US-09-082-253-7
14	163	41.6	95	5	PCT-US96-08235-7
15	61.5	15.7	1257	4	US-08-947-823-5
16	59.5	15.2	114	1	US-08-031-399-3
17	59.5	15.2	114	1	US-08-031-399-6
18	59.5	15.2	114	1	US-08-031-399-12
19	59.5	15.2	114	1	US-08-393-305-3
20	59.5	15.2	114	1	US-08-393-305-6
21	59.5	15.2	114	1	US-08-726-817-3
22	59.5	15.2	114	1	US-08-726-817-6
23	59.5	15.2	114	1	US-08-504-042-3
24	59.5	15.2	114	1	US-08-504-042-6
25	59.5	15.2	114	1	US-08-504-042-12
26	59.5	15.2	114	1	US-08-725-969-3
27	59.5	15.2	114	2	US-08-725-969-6

28	59.5	15.2	114	2	US-08-794-524-3	Sequence 3, Appli
29	59.5	15.2	114	2	US-08-794-524-6	Sequence 6, Appli
30	59.5	15.2	114	4	US-09-189-193-3	Sequence 3, Appli
31	59.5	15.2	114	4	US-09-189-193-6	Sequence 6, Appli
32	59.5	15.2	114	5	PCT-US94-03793-3	Sequence 3, Appli
33	59.5	15.2	114	5	PCT-US94-03793-6	Sequence 6, Appli
34	59.5	15.2	114	5	PCT-US94-03793-12	Sequence 12, Appli
35	59.5	15.2	122	1	US-08-300-903A-3	Sequence 3, Appli
36	59.5	15.2	162	1	US-08-031-399-2	Sequence 2, Appli
37	59.5	15.2	162	1	US-08-031-399-5	Sequence 5, Appli
38	59.5	15.2	162	1	US-08-393-305-2	Sequence 2, Appli
39	59.5	15.2	162	1	US-08-393-305-5	Sequence 5, Appli
40	59.5	15.2	162	1	US-08-535-733-2	Sequence 2, Appli
41	59.5	15.2	162	1	US-08-284-393B-9	Sequence 9, Appli
42	59.5	15.2	162	1	US-08-726-817-2	Sequence 2, Appli
43	59.5	15.2	162	1	US-08-726-817-5	Sequence 5, Appli
44	59.5	15.2	162	1	US-08-504-042-2	Sequence 2, Appli
45	59.5	15.2	162	1	US-08-504-042-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-08-821-451A-6
; Sequence 6, Application US/08821451A
; Patent No. 6066724
; GENERAL INFORMATION:
; APPLICANT: Jian NI, Guo-Liang Yu and Reiner Gentz
; TITLE OF INVENTION: Human Endometrial Specific Steroid-Binding Factor I, II and III
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/821,451A
; FILING DATE: March 21, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/014,724
; FILING DATE: March 21, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: MULLINS, J.G.
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 325800-521 (PF257)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-08-821-451A-6

Query Match 99.5% Score 390; DB 3; Length 95;
Best Local Similarity 98.6%; Pred. No. 5.3e-43;
Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGCKLLEDVTEKINSISIPYKELLOEFIDSDAAAEAMGKQCFLNQSHRTLNKFG 60
Db 19 DSGCKLLEDVTEKINSISIPYKELLOEFIDSDAAAEAMGKQCFLNQSHRTLNKFG 78
QY 61 LMHTVYDYSIWCNL 74
Db 79 LMHTVYDYSIWCNM 92

RESULT 2
US-09-263-810-6
Sequence 6, Application US/09263810
Patent No. 6174992
GENERAL INFORMATION:
APPLICANT: Jian NI, Guo-liang Yu and Reiner Gentz
TITLE OF INVENTION: Human Endometrial Specific Steroid-
TITLE OF INVENTION: Binding Factor I, II and III
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/263.810
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/821.451
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: MULLINS, J.G.
REGISTRATION NUMBER: 33,073
REFERENCE/DOCKET NUMBER: 325800-521 (PF257)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 95 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS:
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-09-263-810-6

Query Match 99.5%; Score 390; DB 4; Length 95;
Best Local Similarity 98.6%; Pred. No. 5.3e-43;
Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 DSGCKLLEDVTEKINSISIPYKELLOEFIDSDAAAEAMGKQCFLNQSHRTLNKFG 60
Db 19 DSGCKLLEDVTEKINSISIPYKELLOEFIDSDAAAEAMGKQCFLNQSHRTLNKFG 78
QY 61 LMHTVYDYSIWCNL 74
Db 79 LMHTVYDYSIWCNM 92

RESULT 3
US-08-455-896-2
Sequence 2, Application US/08455896
Patent No. 568267
GENERAL INFORMATION:
APPLICANT: WATSON, MARK A.
TITLE OF INVENTION: MAMMARY SPECIFIC BREAST CANCER PROTEIN
TITLE OF INVENTION: MAMMARY SPECIFIC BREAST CANCER PROTEIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOWELL & HAFERKAMP, L.C.
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: USA
ZIP: 63105-1817
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

APPLICANT: WATSON, MARK A.
APPLICANT: FLEMING, TIMOTHY P.
TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: USA
ZIP: 63105-1817
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455.896
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: HOLLAND, DONALD R.
REGISTRATION NUMBER: 35,197
REFERENCE/DOCKET NUMBER: 952726
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 727-5188
TELEFAX: (314) 727-6092
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-455-896-2

Query Match 53.1%; Score 208; DB 1; Length 93;
Best Local Similarity 53.4%; Pred. No. 1.2e-19;
Matches 39; Conservative 12; Mismatches 22; Indels 0; Gaps 0;
QY 2 SGCKLLEDVTEKINSISIPYKELLOEFIDSDAAAEAMGKQCFLNQSHRTLNKFG 61
Db 20 SGCKLLEDVTEKINSISIPYKELLOEFIDSDAAAEAMGKQCFLNQSHRTLNKFG 79
QY 62 NMHTVYDYSIWCNL 74
Db 80 FMOLIYDSSLCDL 92

RESULT 4
US-08-933-149-2
Sequence 2, Application US/08933149
Patent No. 5922836
GENERAL INFORMATION:
APPLICANT: WATSON, MARK A.
TITLE OF INVENTION: MAMMARY SPECIFIC BREAST CANCER PROTEIN
TITLE OF INVENTION: MAMMARY SPECIFIC BREAST CANCER PROTEIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOWELL & HAFERKAMP, L.C.
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: USA
ZIP: 63105-1817
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,149
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: HENDERSON, MELODIE W.
REGISTRATION NUMBER: 37,848
REFERENCE/DOCKET NUMBER: 6029-6040
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 727-5188
TELEFAX: (314) 727-6092
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
-08-933-149-2

```

Query Match      53.1%  Score 208;  DB 2;  Length 93;
Best Local Similarity 53.4%  Pred. NO. 1,2e-19;
Matches 39;  Conservative 12;  Mismatches 22;  Indels 0;  Gaps 0;

2  SGCKLLEDMEVKTNSDISIPYKELLQEFIDSDAAAEAMGKFQCFLNQSHRTLKNFGL 61
   ||| |||:: |||| : ||||| ||||| : || : |||||: ||| :
20  SGCPLEENVISKTINPQVSKTEYKELLQEFIDDNATNAIDELKECFLNQDTLSNVEV 79

62  MMHTVYDSIWCNL 74
   | : ||| : ||
80  FMQLIYDSSLQDL 92

```

RESULT 5
 -09-082-343-2
 Sequence 2, Application US/09082343
 Patent No. 5968754
 GENERAL INFORMATION:
 APPLICANT: WATSON, MARK A.
 APPLICANT: FLEMING, TIMOTHY P.
 TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
 TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
 NUMBER OF INVENTIONS: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
 STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
 CITY: ST. LOUIS
 STATE: MISSOURI
 COUNTRY: USA
 ZIP: 63105-1817
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/082.343
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/455,896
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: HOLLAND, DONALD R.
 REGISTRATION NUMBER: 35,197
 REFERENCE/DOCKET NUMBER: 952726
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (314) 727-5188
 TELEFAX: (314) 727-6092
 INFORMATION FOR SEQ ID NO: 2:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
US-09-082-343-2

Query Match 53.1%; Score 208; DB 2; Length 93;
Beat Local Similarity 53.4%; Pred. NO. 1.2e-19;
Matches 39; Conservative 12; Mismatches 22; Indels 0; Gaps 0
QY 2 SGCKLEDMVSKTINSISIPYKELQEEIDSDAAAAMGKFKOCFLNQSHTLKNFL 61
| | | | | : | | | | | | | | | | : | | | | | : | | | | |
Db 20 SCCLPLENVISKTNPQVSKTEYKELQEEIDDNATNAIDELKECFLNQDTLSNVEV 79
| | | | | : | | | | | | | | | | : | | | | | : | | | | |
QY 62 MMHTVYDYSIMCNL 74
| : | | | | : | |
Db 80 FMQLIYDSSLQDL 92

```

```

RESULT 6
US-09-082-253-2
; Sequence 2, Application US/09082253
; Patent No. 6004756
; GENERAL INFORMATION:
; APPLICANT: WATSON, MARK A.
; APPLICANT: FLEMING, TIMOTHY P.
; TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
; TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROGERS, HOWELL & HAFFERKAMP
; STREET: 7733 FORTSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 63105-1817
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/082,253
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/455,896
; FILING DATE: 05/31/1995
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 952726
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; US-09-082-253-2

```

Query Match 53.1%; Score 208; DB 3; Length 93;
Best Local Similarity 53.4%; Pred. No. 1.2e-19;
Matches 39; Conservative 12; Mismatches 22; Indels 0; Gaps 0;

Qy	2	SGCKLEDMVEKTINSIPYEKELLQEFIDSDRAAEAMGFKOCFLNQSHRTLKNFGL	61
		:: : : : :	:
Ddb	20	SGCPULENENKTINPOVSTXEYKELLQEFIDDTATNTNAIDELKECFLNQDTELSNVEV	79
Qy	62	MMHTYDYSINCNL	74
		: :	:
Dd	80	FMOLIYDSSLCDL	92

RESULT 7
 101-US96-08235-2
 Sequence 2. Application PC/TUS9608235

DT-US96-08235-2
 Sequence 2, Application PC/TUS9608235
 GENERAL INFORMATION:
 APPLICANT: WATSON MARK A.

APPLICANT: WATSON, MARK A.
APPLICANT: FLEMING, TIMOTHY P.
TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN

from

TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:

NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROGERS, HOWELL & HAERKAMP
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400

ADDRESS: ROGERS, HOWELL & HAFERKAMP
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI

CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: USA
ZIP: 63105-1817

COUNTRY: USA
ZIP: 63105-1817
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

093
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

SOFTWARE: PatentIn Release #1.0, version #1.23
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/08235
FILING DATE:

009
APPLICATION NUMBER: PC1705360/06233
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: HOLLAND, DONALD R.
REGISTRATION NUMBER: 35,197

NAME: HOLLAND, DONALD R.
REGISTRATION NUMBER: 35,197
REFERENCE/DOCKET NUMBER: 964796
TELECOMMUNICATION INFORMATION:

REFERENCE/DOCREF NUMBER: 304730
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 727-5188
TELEFAX: (314) 727-6092

TELEPHONE: (314) 727-3100
TELEFAX: (314) 727-6092
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid

LENGIN, 35 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

200
SIGNALNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO

000
HYPOTHETICAL: NO
FBI-US96-08235-2

4	0350 00255 2	53.1%; Score 208; DB 5; Length 93;
	Query Match	

Query Match	53.1%	Score 208;	DB 5;	Length 93;
Best Local Similarity	53.4%	Pred. No. 1.2e-19;		
Matches	39; Conservative	12; Mismatches	22; Indels	0; Gaps

	Matches	39;	Conservative	12;	Mismatches	22;	Indels	0;	Gaps	0;
QY	2	SGCKLLDWMVEKTI	NSDISIPEYKELLQE	FIDSDAAAFAMGKF	KQCFLNQSHRTLKNFGL	6				

Qy	2	SGCKLLEDVWKTINS	DISIPEYKELLQEF	TDSDAAAEAMGKFKQCF	LNOSHRTLKNEGL	6
		:: :	:	: :: :	: :: :	
Db	20	SGCPLEENVISKTIN	PQVSKTEYKELLQEF	IDDNATTAIDELKECF	LNOTDETLSNVEV	7

Db	20	SGCPLEUVISKTINPOVSKTEYKELLQEFIDDNATNAIDELKECFLNQTDLSNVEV 7
Qy	62	MMHTVYDSIWCNL 74

Qy	62 MMHTVYDSIWCNL 74
	1. :111 1:1
Db	80 FMQLIYDSSLCDL 92

Db 80 FMQLYDSSLCDL 92

RESULT 8
US-08-821-451A-27

US-08-821-451A-27
; Sequence 27, Application US/08821451A
; Patent No. 6066724

sequence 27, Application 05/00022301;
Patent No. 6066724
GENERAL INFORMATION:
APPLICANT: Jian Ni, Guo-Liang Yu and Reiner Gentz

GENERAL INVENTOR:
; APPLICANT: Jian Ni, Guo-Liang Yu and Relner Gentz
; TITLE OF INVENTION: "Human Endometrial Specific Steroid-

THE EFFECT OF THE INDEPENDENT VARIABLE ON THE DEPENDENT VARIABLE


```

; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/263,810
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/821,451
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: MULLINS, J.G.
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 325800-521 (PF257)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-09-263-810-27

Query Match 41.8%; Score 164; DB 4; Length 95;
Best Local Similarity 36.6%; Pred. No. 5.5e-14;
Matches 26; Conservative 26; Mismatches 19; Indels 0; Gaps 0;

Qy 2 SGCKLLEDVMEKTNISDISIPEYKELLOEFIDSDAAAEAMGKFKOCFLNQSHRILKNFGL 61
Db 21 SGCSILDEVIRGTINSTVTLHDYMKLVKPYQVQAHFTEKAVKQFKOCFLDQDTKLTENVGV 80
Qy 62 MMHTVYDSIWC 72
Db 81 MMEAFNSESC 91

RESULT 10
US-08-455-896-7
Sequence 7, Application US/08455896
Patent No. 5668267
GENERAL INFORMATION:
APPLICANT: WATSON, MARK A.
APPLICANT: FLEMING, TIMOTHY P.
TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: USA
ZIP: 63105-1817
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455,896
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: HOLLAND, DONALD R.
REGISTRATION NUMBER: 35,197
REFERENCE/DOCKET NUMBER: 952726
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 727-5188
TELEFAX: (314) 727-6092
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 95 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-455-896-7

Query Match 41.8%; Score 164; DB 4; Length 95;
Best Local Similarity 36.6%; Pred. No. 5.5e-14;
Matches 26; Conservative 26; Mismatches 19; Indels 0; Gaps 0;

Qy 2 SGCKLLEDVMEKTNISDISIPEYKELLOEFIDSDAAAEAMGKFKOCFLNQSHRILKNFGL 61
Db 21 SGCSILDEVIRGTINSTVTLHDYMKLVKPYQVQAHFTEKAVKQFKOCFLDQDTKLTENVGV 80
Qy 62 MMHTVYDSIWC 72
Db 81 MMEAFNSESC 91

RESULT 11
US-08-933-149-7
Sequence 7, Application US/08933149
Patent No. 5922836
GENERAL INFORMATION:
APPLICANT: WATSON, MARK A.
APPLICANT: FLEMING, TIMOTHY P.
TITLE OF INVENTION: MAMMAGLOBIN, A SECRETED
TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOWELL & HAFERKAMP, L.C.
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: USA
ZIP: 63105-1817
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,149
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: HENDERSON, MELODIE W.
REGISTRATION NUMBER: 37,848
REFERENCE/DOCKET NUMBER: 6029-6040
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 727-5188
TELEFAX: (314) 727-6092
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 95 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-933-149-7

Query Match 41.6%; Score 163; DB 2; Length 95;
Best Local Similarity 36.6%; Pred. No. 7.4e-14;
Matches 26; Conservative 25; Mismatches 20; Indels 0; Gaps 0;

Qy 2 SGCKLLEDVMEKTNISDISIPEYKELLOEFIDSDAAAEAMGKFKOCFLNQSHRILKNFGL 61
Db 21 SGCSILDEVIRGTINSTVTLHDYMKLVKPYQVQAHFTEKAVKQFKOCFLDQDTKLTENVGV 80
Qy 62 MMHTVYDSIWC 72
Db 81 MMEAFNSESC 91
```

```

;
; TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
; TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 63105-1817
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/082,253
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/455,896
; FILING DATE: 05/31/1995
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 952726
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-09-082-253-7

Query Match 41.6%; Score 163; DB 3; Length 95;
Best Local Similarity 36.6%; Pred. No. 7.4e-14;
Matches 26; Conservative 25; Mismatches 20; Indels 0; Gaps 0;

QY 2 SGCKLEDWVKTINSIDISPEYKELLQEFIDSDAAAEAMGFKQCFLNQSHRTLKNFGL 61
Db 21 SGCSILDEVIRGTINSTVTLHDYMKLVKPYQDHFTEKAVKQFKCFLDQDTKLTENVGV 80
QY 62 MMHTYDYSWC 72
Db 81 MMEALFNSESC 91

RESULT 14
PCT-US96-08235-7
; Sequence 7, Application PC/TUS9608235
; GENERAL INFORMATION:
; APPLICANT: WATSON, MARK A.
; APPLICANT: FLEMING, TIMOTHY P.
; TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
; TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 63105-1817
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; APPLICANT: FLEMING, TIMOTHY P.

;
; TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
; TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 63105-1817
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/082,343
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/455,896
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 952726
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; 09-082-343-7

Query Match 41.6%; Score 163; DB 2; Length 95;
Best Local Similarity 36.6%; Pred. No. 7.4e-14;
Matches 26; Conservative 25; Mismatches 20; Indels 0; Gaps 0;

QY 2 SGCKLEDWVKTINSIDISPEYKELLQEFIDSDAAAEAMGFKQCFLNQSHRTLKNFGL 61
Db 21 SGCSILDEVIRGTINSTVTLHDYMKLVKPYQDHFTEKAVKQFKCFLDQDTKLTENVGV 80
QY 62 MMHTYDYSWC 72
Db 81 MMEALFNSESC 91

RESULT 13
US-09-082-253-7
; Sequence 7, Application US/09082253
; Patent No. 5968754
; GENERAL INFORMATION:
; APPLICANT: WATSON, MARK A.
; APPLICANT: FLEMING, TIMOTHY P.

```

Copied from PC 009367009 on 27-04-2004

SERIAL 15
 -08-947-823-5
 Sequence 5, Application US/08947823
 Patent No. 6114605
 GENERAL INFORMATION:
 APPLICANT: Williamson, Valerie M.
 APPLICANT: Kadohian, Isghouhi
 APPLICANT: Yaghoobi, Jafar
 APPLICANT: Bodeau, John
 APPLICANT: Milligan, Stephen
 TITLE OF INVENTION: Procedures and Materials for Conferring
 TITLE OF INVENTION: Pest Resistance in Plants
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/947,823
 FILING DATE: 09-OCT-1997
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 PRIOR APPLICATION NUMBER: PCT/US97/18802
 FILING DATE: 09-OCT-1997
 PRIOR APPLICATION NUMBER: US 60/028,191
 FILING DATE: 10-OCT-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Bastiad, Kevin L.

```

; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 023070-070210US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1257 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-947-823-5

```

us-09-367-009-3.rai

Tue May 1 11:45:25 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 15:11:53 ; Search time 20.75 Seconds
(without alignments)
203.859 Million cell updates/sec

Title: US-09-367-009-3
Perfect score: 392
Sequence: 1 DSGCKLLEDVMEKTVINSIS.....TLKNFGLMHTVYDSIWCNL 74

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

- Database :
- 1: A_Geneseq_0401.*
 - 2: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
 - 3: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
 - 4: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
 - 5: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
 - 6: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
 - 7: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
 - 8: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
 - 9: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
 - 10: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
 - 11: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
 - 12: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
 - 13: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
 - 14: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
 - 15: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
 - 16: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
 - 17: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
 - 18: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
 - 19: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
 - 20: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
 - 21: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
 - 22: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*
 - 23: /SIDSL/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	392	100.0	74	19 W61649	Non-ocular disease
2	390	99.5	95	18 W35804	Human endometrial
3	390	99.5	95	20 Y02590	A human mamoglobi
4	390	99.5	95	21 B03769	Human endometrial
5	390	99.5	95	21 Y92226	Human endometrial
6	390	99.5	95	21 Y92237	Mamoglobin homolo
7	390	99.5	95	21 Y65394	Human 5' EST relat
8	390	99.5	108	20 Y60038	Human endometrium
9	208	53.1	74	21 Y84624	Amino acid sequenc
10	208	53.1	93	18 W10179	Mammary-specific s
11	208	53.1	93	19 W59777	Amino acid sequenc

12	208	53.1	93	19 W48432	Mamoglobin protei
13	208	53.1	93	20 Y01718	Mamoglobin, a mam
14	208	53.1	93	21 Y84622	Amino acid sequenc
15	208	53.1	93	22 B51127	Human mamoglobin
16	95	24.2	33	19 W48433	Mamoglobin synthe
17	79	20.2	30	19 W48435	Mamoglobin synthe
18	72	18.4	20	22 B51132	Human mamoglobin
19	66	16.8	20	22 B51115	Human mamoglobin
20	66	16.8	20	22 B51133	Human mamoglobin
21	65	16.6	22	19 W48441	Mamoglobin synthe
22	64.5	16.5	395	21 G48524	Arabidopsis thalia
23	64.5	16.5	446	21 G48523	Arabidopsis thalia
24	64.5	16.5	610	21 G48522	Arabidopsis thalia
25	64.5	16.5	1002	9 P81861	Sequence encoded b
26	62.5	15.9	2485	21 B18172	Plasmodium falcipa
27	62	15.8	27	19 W48434	Mamoglobin synthe
28	62	15.8	764	21 B18282	Plasmodium falcipa
29	62	15.8	2013	21 B18265	Plasmodium falcipa
30	61.5	15.7	1206	19 W47080	Tomato MI resistan
31	61.5	15.7	1206	19 W53583	Wild tomato MI res
32	61.5	15.7	1257	19 W53575	Tomato pest resist
33	61.5	15.7	1257	19 W53582	Wild tomato MI res
34	61	15.6	732	19 W40496	Mouse XRCC5 protei
35	60.5	15.4	325	21 G26324	Arabidopsis thalia
36	60.5	15.4	325	21 G26324	Arabidopsis thalia
37	60.5	15.4	328	21 G26323	Arabidopsis thalia
38	60.5	15.4	328	21 G26322	Arabidopsis thalia
39	60.5	15.4	430	21 G26322	Arabidopsis thalia
40	60.5	15.4	430	21 G48079	Arabidopsis thalia
41	59.5	15.2	114	16 R83310	Human interleukin
42	59.5	15.2	114	16 R83310	Human interleukin
43	59.5	15.2	114	16 R83435	Mammalian interlev
44	59.5	15.2	114	16 R65928	Mammalian IL-15.
45	59.5	15.2	114	17 W69099	Simian mature epit

ALIGNMENTS

RESULT 1

- W61649
- ID W61649 standard; peptide: 74 AA.
- XX
- AC W61649;
- XX
- DT 27-OCT-1998 (first entry)
- DE
- XX Non-ocular disease marker 3.
- XX
- KW Human; non-ocular disease; tear; cancer; breast; prostate.
- XX
- OS Homo sapiens.
- XX
- PN W09835229-A1.
- XX
- PD 13-AUG-1998.
- XX
- PF 06-FEB-1998; 98WO-AU000071.
- XX
- PR 07-FEB-1997; 97AU-0005009.
- XX
- PA (MACQ-) MACQUARIE RES LTD.
- XX (UNIX) UNISEARCH LTD.
- PA
- XX
- PI Bolis S, Gooley AA, Herbert B, Molloy M, Morris C;
- PI Walsh B, Willcox M, Williams KL;
- XX
- DR WPI; 1998-447373/38.
- XX
- PT Screening for non-ocular disease - by analysing tears for marker
- PT proteins, particularly indicative of cancer and genetic disease,
- PT also new proteins and nucleic acid encoding them
- XX

PS Claim 6; Page 9; 14pp; English.
XX The markers W61647-W61649 are used for screening for, or detecting,
CC non-ocular disease by analysing tears. Biochemicals, specifically
CC proteins, are isolated from tears, particularly by chromatography or
CC electrophoresis, especially two-dimensional polyacrylamide gel
CC electrophoresis (2D-PAGE), then detected, e.g. with labelled specific
CC reagents, in (radio)immunoassay. The method is used to detect cancer,
CC particularly of breast or prostate, or a genetic disease, in humans or
CC animals.

XX Sequence 74 AA;
XX Query Match 100.0%; Score 392; DB 19; Length 74;
XX Best Local Similarity 100.0%; Pred. No. 3.8e-41;
XX Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 DSGCKLLEDVVEKTSINSDISIPYKELQLQEFIDSDAAEAMGKFKQCFLNQSHRTLNFG 60
|||||
1 dsgcklledmvektinsdisipyeikellqefidsdaaaamgkfkqcfngshrtlnkf 60
61 LMHTVYDSIWCNL 74
|||||
61 lmhtvydsiwcnl 74

RESULT 2
W35804
W35804 standard; Protein; 95 AA.
W35804;

27-MAR-1998 (first entry)
Human endometrial specific steroid-binding factor III.
Endometrial specific steroid-binding factor III; ESF III; human;
Clara cell secretory protein; endometrium;
phospholipase A2 inhibitor; polychlorinated biphenyl; antiaggregant;
inflammation; asthma; rhinitis; cystic fibrosis; airway disease;
neoplasia; atopy; therapy; diagnosis.

Homo sapiens.
Key Location/Qualifiers
Peptide 1..21
Protein 22..95
/label= Sig_peptide
/label= Mat_protein

WO9734997-A1.
25-SEP-1997.
21-MAR-1996; 96WO-US03857.
21-MAR-1996; 96WO-US03857.
(HUMA-) HUMAN GENOME SCI INC.
Gentz RL, Ni J, Yu G;
WPI; 1997-480206/44.
N-PSDB; T94832.

Human endometrial specific steroid-binding factor I, II and III -
used to treat inflammation, asthma, rhinitis, cystic fibrosis,
airway disease, neoplasia, atopy etc.
Claim 19; Page 65; 92pp; English.
This sequence comprises human endometrial specific steroid binding

CC factor III (ESF III), a protein that inhibits phospholipase A2
CC activity, binds to polychlorinated biphenyl compounds, reduces
CC foreign protein antigenicity, inhibits monocyte and neutrophil
CC chemotaxis and phagocytosis, inhibits platelet aggregation,
CC regulates eicosanoid levels in the human uterus and controls the
CC growth of endometrial cells. The amino acid sequence was deduced
CC from a cDNA clone (see T94832) derived from a human endometrial
CC tumour. ESF I (see W35802) and ESF II (see W35803) are also
CC claimed. Human ESF III has about 36% identity with rat prostatic
CC steroid-binding protein. Recombinant ESF I, II and III can be
CC expressed in host cells for use in claimed methods (a) for treating
CC a patient in need of ESF I, II or III (including expression of the
CC polypeptide in vivo) and (b) for identifying compounds which bind
CC to and inhibit activation of the ESF polypeptide. ESF I, II and
CC III may be used to treat inflammation, asthma, rhinitis, cystic
CC fibrosis, airway disease, neoplasia and atopy.

XX Sequence 95 AA;

Query Match 99.5%; Score 390; DB 18; Length 95;
Best Local Similarity 98.6%; Pred. No. 9.3e-41;
Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 DSGCKLLEDVVEKTSINSDISIPYKELQLQEFIDSDAAEAMGKFKQCFLNQSHRTLNFG 60
Db 19 dsgcklledmvektinsdisipyeikellqefidsdaaaamgkfkqcfngshrtlnkf 78
QY 61 LMHTVYDSIWCNL 74
Db 79 lmhtvydsiwcnm 92

RESULT 3
Y02590
ID Y02590 standard; Protein; 95 AA.
XX AC Y02590;
XX 26-JUL-1999 (first entry)
XX A human mammoglobin homologue (HMH).
XX Human mammoglobin homologue; HMH; antagonist; neoplastic disorder;
XX adenocarcinoma; leukemia; lymphoma; melanoma; myeloma; sarcoma;
XX teratocarcinoma; endometriosis.
XX Homo sapiens.
XX WO9919487-A1.
XX 22-APR-1999.
XX 14-OCT-1998; 98WO-US21729.
XX 16-OCT-1997; 97US-0951750.
XX (INCY-) INCYTE PHARM INC.
XX Hillman JL, Murry LE, Shah P;
XX WPI; 1999-302531/25.
XX N-PSDB; X36138.
XX New human mammoglobin homologue (HMH), useful for diagnosing, treating
XX or preventing disorders associated with expression of HMH
XX Claim 1; Fig 1A-B; 63pp; English.
XX The present sequence represents a human mammoglobin homologue (HMH).
XX Antagonists of the HMH polypeptide can be used to treat neoplastic
XX disorders including adenocarcinoma, leukemia, lymphoma, melanoma,
XX myeloma, sarcoma and teratocarcinoma. A vector expressing the

complement of the polynucleotide encoding HMH may be administered to a subject to treat or prevent neoplastic disorders or endometriosis. Antibodies which bind HMH may also be used in the diagnosis of conditions or diseases characterized by expression of HMH, or in assays to monitor patients being treated with HMH, agonists, antagonists or inhibitors. Polynucleotides encoding HMH may also be used diagnostically to detect and quantitate gene expression in biopsied tissues. With respect to cancer a relatively high amount of transcript may indicate a predisposition for the development of disease. The nucleic acid sequences which encode HMH may also be used to generate hybridization probes useful for mapping the naturally occurring genomic sequence. HMH, and its fragments/variants can be used for screening libraries of compounds in drug screening techniques.

Sequence 95 AA;

Query Match 99.5%; Score 390; DB 20; Length 95;
Best Local Similarity 98.6%; Pred. No. 9.3e-41;
Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 DSGCKLLEDVMEKTSINSIPYKELQEFIDSDAAAMGKFKQCFNLNOSHRITLKNFG 60
|||||
19 dsgcklledvmektinsdisipeykellqefidsdaaamgkfkqcfngshrtlkknfg 78

61 LMMHTVYDSIWCNL 74
|||||
79 lmmhtvydsiwcnm 92

RESULT 4

B03769 standard; Protein; 95 AA.

B03769;

06-OCT-2000 (first entry)

Human endometrial specific steroid-binding factor III protein sequence.
Endometrial specific steroid-binding factor; human; hESF; inflammation;
asthma; rhinitis; cystic fibrosis; air way disease; neoplasia; atopy;
eicosanoid level regulator; chemotaxis inhibitor; endometrial cancer; ss.

Homo sapiens.

US6066724-A.

23-MAY-2000.

21-MAR-1997; 97US-0821451.

21-MAR-1996; 96US-0014724.

(HUMA-) HUMAN GENOME SCI INC.

Yu G, Gentz R, N1 J;

WPI; 2000-375600/32.

N-PSDB; A59730.

Novel gene encoding human endometrial specific steroid-binding factor I, II and III which is useful for treating asthma, rhinitis, cystic fibrosis, airway disease and neoplasia

Claim 1; Fig 3; 36pp; English.

This invention relates to nucleic acid molecules encoding portions of the human endometrial specific steroid-binding factors I, II, and III. Also included in the invention are hESF I, II, and III polypeptide sequences. The nucleotide sequence exhibit antiasthmatic, antiinflammatory, antiallergic, and cytostatic properties. The polynucleotides are used in gene therapy to express hESF I, II and III polypeptides in vivo to treat

and/or prevent inflammation, asthma, rhinitis, cystic fibrosis, air way disease, neoplasia and atopy. The polynucleotides are also used to inhibit phospholipase A2 activity, bind polychlorinated biphenyls, reduce foreign protein antigenicity, inhibit monocyte and neutrophil chemotaxis and phagocytosis, inhibit platelet aggregation, regulate eicosanoid levels in the human uterus and control the growth of endometrial cells. The polynucleotides are also useful for detecting complementary polynucleotides as a diagnostic reagent. The hESF I, II and III polynucleotides are used to detect complementary polynucleotides such as a diagnostic reagent. Detection of a mutated form of hESF I, II and III associated with a dysfunction will provide a diagnostic tool that can define diagnosis of a disease or susceptibility to a disease which results from under-expression, over-expression or altered expression of hESF I, II and III e.g. a susceptibility to inherited asthma and endometrial cancer. They are also useful for chromosome identification. The present sequence represents a hESF III protein sequence identified in the invention.

Sequence 95 AA;

Query Match 99.5%; Score 390; DB 21; Length 95;
Best Local Similarity 98.6%; Pred. No. 9.3e-41;
Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 DSGCKLLEDVMEKTSINSIPYKELQEFIDSDAAAMGKFKQCFNLNOSHRITLKNFG 60
|||||
19 dsgcklledvmektinsdisipeykellqefidsdaaamgkfkqcfngshrtlkknfg 78

61 LMMHTVYDSIWCNL 74

|||||

79 lmmhtvydsiwcnm 92

RESULT 5

Y92226

ID Y92226 standard; Protein; 95 AA.

XX AC Y92226;

XX DT 10-AUG-2000 (first entry)

XX DE Human endometrial specific steroid binding factor III.

XX Endometrial specific steroid binding factor; ESBPIII; diagnosis;
XX gynaecological cancer; uterine; breast; endometrial; ovarian;
XX antibody; gene therapy; cytostatic.

XX OS Homo sapiens.

XX WO200020044-A1.

XX PD 13-APR-2000.

XX 30-SEP-1999; 99WO-US222753.

XX 02-OCT-1998; 98US-0102743.

XX (DIAD-) DIADEXUS LLC.

XX Macina RA;

XX WPI; 2000-303649/26.

XX N-PSDB; A09104.

XX Diagnosing, staging and monitoring gynecological cancer comprising using an elevated level of ESBPIII in a patient as an indicator of cancer

XX Claim 6; Page 28-29; 32pp; English.

XX The levels of human endometrial specific steroid binding factor (ESBPIII) can be measured and compared to control levels and used to

CC diagnose the presence of a gynaecological (uterine, breast, endometrial,
CC or ovarian) cancer in a patient. ESBPIII levels can also be used to
CC diagnose metastasis, to stage or monitor gynaecological cancer.
CC Antibodies specific for ESBPIII can be used to treat gynecological
CC cancers.
XX
SQ Sequence 95 AA;

Query Match 99.5%; Score 390; DB 21; Length 95;
Best Local Similarity 98.6%; Pred. No. 9.3e-41;
Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 DSGCKLLEDVVEKTSINSDISPEYKELLOQFIDSDAAEAMGKFCQFLNQSHRTLKNFG 60
Db 19 dsgcklledmvektinsdispeykellqefidsdaaaeamgkfkqcfingshrtlknfg 78
QY 61 LMHTVYDYSIWCNL 74
Db 79 lmmhtvydyswcnm 92

RESULT 6
Y92237 Y92237 standard; Protein; 95 AA.
Y92237;
10-AUG-2000 (first entry)
Mammoglobin homologue from clone Mamm-X.
Clone Mamm-X; mammaglobin; breast cancer; cytostatic; anti-HIV;
immunosuppressive; antiallergic; antineoplastic; antiinflammatory;
antiarthritic; antiarteriosclerotic; vasotropic; neuroprotective;
nootropic; dermatological; tranquilizer; vulnerary.

Homo sapiens.
WO200020447-A2.
13-APR-2000.
06-OCT-1999; 99WO-US23294.
06-OCT-1999; 98US-0103195.
05-OCT-1999; 99US-0103195.
(CURA-) CURAGEN CORP.
Shinkets RA;
WPI; 2000-303741/26.
N-PSDB; A09118.
Nucleic acids encoding polypeptides with syncollin-like, claudin-like or
cytokine-like activity, useful for treating diseases including cancer,
PT Alzheimer's and atherosclerosis
PS Claim 23; Fig 6; 118pp; English.
XX Clone Mamm-X encodes a polypeptide that is 100 percent identical to human
CC Mammaglobin B precursor, a potential marker of breast cancer nodal
CC metastasis. The sequences are useful for treatment of diseases such as
CC cancer, immune disorders, autoimmune disease, acquired immune deficiency
CC syndrome (AIDS), transplant rejection, allergy, infection by a
CC pathological agent or organism, inflammatory disorders, arthritis, a
CC hematopoietic disorder, a skin disorder, atherosclerosis, restenosis, a
CC neurological disease, Alzheimer's disease, trauma, spinal cord injury
CC and skeletal disorders.
XX Sequence 95 AA;
SQ

Query Match 99.5%; Score 390; DB 21; Length 95;
Best Local Similarity 98.6%; Pred. No. 9.3e-41;
Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 DSGCKLLEDVVEKTSINSDISPEYKELLOQFIDSDAAEAMGKFCQFLNQSHRTLKNFG 60
Db 19 dsgcklledmvektinsdispeykellqefidsdaaaeamgkfkqcfingshrtlknfg 78
QY 61 LMHTVYDYSIWCNL 74
Db 79 lmmhtvydyswcnm 92

RESULT 7
Y65394 Y65394 standard; Protein; 95 AA.
XX Y65394;
XX 01-FEB-2000 (first entry)
XX Human 5' EST related polypeptide SEQ ID NO:1555.
XX Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW gene therapy; chromosome mapping; upstream regulatory sequence;
KW forensic; location; development; protein synthesis; stability;
KW regulation; identification.
XX Homo sapiens.
OS WO9953051-A2.
XX 21-OCT-1999.
XX 09-APR-1999; 99WO-IB00712.
XX 09-APR-1998; 98US-0057719.
XX 28-APR-1998; 98US-0069047.
XX (GEST) GENSET.
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI WPI; 2000-038446/03.
XX N-PSDB; 243008.
PT Novel secreted protein 5' expressed sequence tag sequences used in
PT diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX Claim 3; Page 818; 837pp; English.
XX 242265 to 243075 represent novel 5' expressed sequence tag (EST)
CC sequences, corresponding to human secreted proteins. Y64651 to Y65438
CC represent the EST-related proteins corresponding to 242265 to 243052.
CC The 5' ESTs can be used for producing secreted human gene products.
CC They can be used to identify and isolate 5' untranslated regions (UTRs)
CC and upstream regulatory regions which control the location, development
CC stage, rate, and quantity of protein synthesis, as well as stability of
CC mRNA. The ESTs are also useful as probes for chromosome mapping, and to
CC obtain full length cDNA clones. The ESTs can also be used in forensic
CC procedures to identify individuals, or in diagnostic procedures to
CC identify individuals having genetic diseases resulting from abnormal
CC gene expression. The products may also be used in gene therapy protocols.
CC The nucleic acids encoding signal peptides can be used for directing
CC extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.
CC The proteins encoded by the EST sequences may be useful in treating a
CC variety of human conditions. Secreted proteins have therapeutic value,
CC and the identification of new secreted proteins is valuable. 242249 to
CC 242264 and Y64644 to Y64650 represent sequences used in the
CC exemplification of the present invention.
XX

Query Match 99.5%; Score 390; DB 20; Length 108;

Mammaglobin; detection; diagnosis; breast cancer; tumour; antibody;
gene therapy; human.

Homo sapiens.

WO9807753-A1.

26-FEB-1998.

19-AUG-1997; 97WO-US14666.

15-AUG-1997; 97US-0697106.

19-AUG-1996; 96US-0697106.

(ABBO) ABBOTT LAB.

Billings-Medel PA, Cohen M, Colpitts TL, Friedman PN;
Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD;
Roberts-Rapp L, Russell JC, Stroupe SD;

WPI: 1998-169096/15.

N-PSDB; V17905, V17906.

Antibodies to mammaglobin polypeptide(s) - used for detecting,
diagnosing, preventing or treating diseases or conditions of breast
such as breast cancer

Claim 8; Page 92; 105pp; English.

The present sequence represents mammaglobin which is used in an example
of the present invention. The present invention describes an antibody (A)
which specifically binds to at least 1 mammaglobin epitope (WE) which is
derived from an amino acid sequence having at least 50% identity to an
amino acid sequence (see W48432) and fragments. Also described are: (1)
an assay kit for determining the presence of mammaglobin antigen (MA) in
a test sample, comprising a container containing an antibody as in (A);
(2) a method for producing antibodies which specifically bind to a MA,
comprising administering an isolated immunogenic polypeptide or fragment
to elicit an immune response, where the immunogenic polypeptide
comprises at least 1 ME and has at least 50% identity to a sequence
(see W48432) and fragments, and (3) a method for producing antibodies
which specifically bind to a MA comprising administering to a mammal a
plasmid comprising a sequence which encodes at least 1 ME derived from a
polypeptide having an amino acid sequence (see W48432) and fragments.
The products and methods can be used for detecting, diagnosing, staging,
monitoring, prognosticating, preventing or treating, or determining
predisposition to diseases or conditions of the breast such as breast
cancer.

Sequence 93 AA;

Query Match 53.1%; Score 208; DB 19; Length 93;
Best Local Similarity 53.4%; Pred. No. 2.5e-18;
Matches 39; Conservative 12; Mismatches 22; Indels 0; Gaps 0;

2 SGCKLEDMVETKINSISIPYKELLOEFIDSDAAAEAMGKFCFLNQSHRTLNKFL 61
||| ||||| : ||| ||||| : ||| ||||| : ||| ||||| : ||| :
20 sgcpllenvisktinpqvskteykellqefidnattnaidkefcflngtdetlsnvev 79

62 MMHTVYDSIWCNL 74

80 fmqllydsldcl 92

RESULT 13

Y01718

ID Y01718 standard; Protein; 93 AA.

AC Y01718;

XX

25-JUN-1999 (first entry)

XX

DE Mammaglobin, a mammary specific protein.
XX
KW Human; mammary-specific protein; mammaglobin; antigen; vaccine;
KW mammaglobin-expressing cancer; breast cancer;
KW autoLogous tumor lymphocyte; diagnosis; marker.
XX
OS Homo sapiens.
XX
PN WO9914230-A1.
XX
PD 25-MAR-1999.
XX
PF 18-SEP-1998; 98WO-US17991.
XX
PR 18-SEP-1997; 97US-0933149.
XX
PA (UNIW) UNIV WASHINGTON.
XX
PI Fleming TP, Watson MA;
XX
DR WPI: 1999-244021/20.
DR N-PSDB; X26966.
XX
Mammaglobin, secreted protein overexpressed in breast cancer
XX
Claim 15; Fig 2; 60pp; English.
XX
The present sequence represents a human mammary-specific protein,
designated mammaglobin. The specification describes a protein
comprising a mammaglobin antigen that is recognized by B and/or
Tc cells specific for the natural, secreted and glycosylated form
of mammaglobin polypeptide. This protein, or recombinant vectors
that express it, are used in vaccines for treating mammaglobin-
expressing cancers, specifically of the breast. Such cancers can
also be treated using autologous tumor lymphocytes activated
ex vivo with an mammaglobin antigen, then returned to the
patient. Expression of mammaglobin is elevated in 27% of stage I
primary breast cancers, so it represents a marker useful for
diagnosis of this disease.
XX
Sequence 93 AA;

Query Match 53.1%; Score 208; DB 20; Length 93;
Best Local Similarity 53.4%; Pred. No. 2.5e-18;
Matches 39; Conservative 12; Mismatches 22; Indels 0; Gaps 0;

2 SGCKLEDMVETKINSISIPYKELLOEFIDSDAAAEAMGKFCFLNQSHRTLNKFL 61
||| ||||| : ||| ||||| : ||| ||||| : ||| ||||| : ||| :
20 sgcpllenvisktinpqvskteykellqefidnattnaidkefcflngtdetlsnvev 79

62 MMHTVYDSIWCNL 74

80 fmqllydsldcl 92

RESULT 14

Y84622

ID Y84622 standard; Protein; 93 AA.

AC Y84622;

XX

25-JUL-2000 (first entry)

XX

Amino acid sequence of the mammary-specific protein mammaglobin.

Human; mammaglobin; mammary gland; breast cancer; endometrial cancer.

OS Homo sapiens.

PN WO200018783-A1.

XX

06-APR-2000.

PD

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: April 24, 2001, 15:12:57 ; Search time 13.62 Seconds
(without alignments)
373.385 Million cell updates/sec

Title: US-09-367-009-3
Perfect score: 392
Sequence: 1 DSGCKLLEDVKEVNTINDIS.....TLKNFGLMHMTVYDSIWCNL 74

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_67.*

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	163	41.6	95	1 BORT3	prostatic steroid-
2	76.5	19.5	280	2 T38449	hypothetical prote
3	70	17.9	303	2 B71638	cell division prot
4	66.5	17.0	234	2 F64104	hypothetical prote
5	65.5	16.7	131	2 F70880	hypothetical prote
6	64.5	16.7	603	1 W1WL8	El protein - human
7	64.5	16.5	153	2 E82928	hypothetical prote
8	64.5	16.5	610	2 S59558	dynam-in-like prote
9	63.5	16.2	269	2 G75148	hypothetical prote
10	63	16.1	131	2 D81908	hypothetical prote
11	62.5	15.9	496	2 T46356	hypothetical prote
12	62.5	15.9	2485	1 H71621	serine/threonine-s
13	62	15.8	131	2 D81108	conserved hypoteth
14	62	15.8	764	2 H71607	hypothetical prote
15	62	15.8	2013	2 C71610	probable membrane
16	61.5	15.7	162	2 I49124	interleukin-15 m
17	61.5	15.7	734	2 E72271	5-methyltetrahydro
18	61.5	15.7	1257	2 T06269	root-knot nematode
19	61	15.6	490	2 T31646	hypothetical prote
20	61	15.6	605	2 S36469	El protein - human
21	61	15.6	663	2 T37772	telomere length re
22	61	15.6	732	2 S26303	Ku autoantigen 80K
23	60.5	15.4	269	2 G71188	hypothetical prote
24	60.5	15.4	430	2 T04668	phosphoserine tran
25	60.5	15.4	606	2 S60090	rad17 protein - fi
26	60.5	15.4	611	2 S61147	TCM10 protein - ye
27	60.5	15.4	1213	2 T41378	probable helicase
28	60.5	15.4	1601	2 T18800	hypothetical prote
29	60	15.3	268	1 GWKECC	indole-3-glycerol-

30	60	15.3	534	2 A37483	F protein - Muraya
31	60	15.3	606	1 W1WL85	El protein - human
32	60	15.3	790	1 WFMF9	90K protein - alfa
33	60	15.3	1071	2 S38164	ATP-binding protei
34	60	15.3	3795	2 T00831	hypothetical prote
35	59.5	15.2	162	1 A53484	interleukin-15 pre
36	59.5	15.2	312	2 E70376	exopolysphatase
37	59.5	15.2	785	2 D71653	cell surface anlig
38	59.5	15.2	925	2 T37831	hypothetical prote
39	59.5	15.2	1075	2 S54067	probable membrane
40	59.5	15.2	1177	2 I64233	hypothetical prote
41	59.5	15.2	1198	2 T20262	hypothetical prote
42	59.5	15.2	1255	2 T06267	nenatodes resistan
43	59	15.1	173	2 T01889	hypothetical prote
44	59	15.1	604	2 S36540	El protein - human
45	59	15.1	606	1 W1WL5	El protein - human

ALIGNMENTS

RESULT 1

BORT3
prostatic steroid-binding protein chain C3 precursor - rat
N:Alternate names: prostatin
C:Species: Rattus norvegicus (Norway rat)
C:Date: 17-Dec-1982 #sequence_revision 14-Nov-1993 #text_change 28-May-1999
C:Accession: A92395; A92433; A91108; A42392; A03250
R:Parker, M.G.; White, R.; Hurst, H.; Needham, M.; Tilly, R.
J Biol. Chem. 258, 12-15, 1983
A:Title: Prostatic steroid-binding protein. Isolation and characterization of C3 gene
A:Reference number: A92395; MUID:83082848
A:Accession: A92395
A:Molecule type: mRNA
A:Residues: 1-95 <PAR>
A:Cross-references: GB:V01263; GB:J00777; NID:q56993; PIDN:CAA24577.1; PID:q56994
R:Viskochil, D.H.; Perry, S.T.; Lea, O.A.; Stafford, D.W.; Wilson, E.M.; French, F.S.
J. Biol. Chem. 258, 8861-8866, 1983
A:Title: Isolation of two genomic sequences encoding the M-r = 14,000 subunit of rat
A:Reference number: A92433; MUID:83238526
A:Accession: A92433
A:Molecule type: mRNA
A:Residues: 1-78, S', 80-95 <VIS>
R:Peeters, B.; Rombauts, W.; Mous, J.; Heyns, W.
Eur. J. Biochem. 115, 115-121, 1981
A:Title: Structural studies on rat prostatic binding protein. The primary structure o
A:Reference number: A91108; MUID:81188769
A:Accession: A91108
A:Molecule type: protein
A:Residues: 19-95 <PEE>
R:Tan, J.A.; Marschke, K.B.; Ho, K.C.; Perry, S.T.; Wilson, E.M.; French, F.S.
J. Biol. Chem. 267, 4456-4466, 1992
A:Title: Response elements of the androgen-regulated C3 gene.
A:Reference number: A42392; MUID:92165796
A:Accession: A42392
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-16, T', 18-95 <TAN>
A:Note: sequence inconsistent with the nucleotide translation
C:Comment: C3 is encoded by two unique genes that differ from each other only in thei

C:Comment: Steroid-binding protein, the principal androgen-dependent secretory protei
ng C2 and C3 chains. The chains of each dimer are linked by disulfide bonds. This pro
C:Superfamily: uteroglobin
C:Keywords: glycoprotein
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-95/Product: prostatic steroid-binding protein C3 chain #status experimental <MAT
F:35/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 41.6%; Score 163; DB 1; Length 95;
Best Local Similarity 36.6%; Pred. No. 7.5e-11;


```

RESULT      6
WLWLB
E1 protein - human papillomavirus type 8
C:Species: human papillomavirus type 8
C>Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 07-Nov-1997
C:Accession: A03856
R:Fuchs, P.G.; Iftner, T.; Weninger, J.; Pfister, H.
J. Virol. 58, 626-634, 1986
A>Title: Epidermodysplasia verruciformis-associated human papillomavirus 8: genomic sequence
A:Reference number: A93019; MUID:86200410
A:Accession: A03656
A:Molecule type: DNA
A:Residues: 1-603 <FUC>
A:Cross-references: GB:M12737; NID:g333074
A>Note: this ORF is not annotated in GenBank entry pPH8CG
C:Superfamily: papillomavirus E1 protein
C:Keywords: early protein

Query Match          16.7%; Score 65.5; DB 1; Length 603;
Best Local Similarity 22.2%; Pred. No. 23;
Matches 22; Conservative 21; Mismatches 23; Indels 33; Gaps 4;

Oy 1 DSGCKLEDVMEKVTINSDISIP-----EYKELQEFIDSAAAEAMGK 43
||| : | : : : : : : : : : : : : : : : : : : : : : :
Db 112 DSGVELTLNNEADVSHEVEPAIDSPEDBGGSGALDIDYTALLR---SSNTKATLMAK 168
||| : | : : : : : : : : : : : : : : : : : : : : : :
Oy 44 FKOCFLN-----QSRLTKNFGLM----MHTVYDS 69
||| : | : : : : : : : : : : : : : : : : : : : : : :
Db 169 FKEAFGDGFNELTRQFSKYKTCCNYVVVAAYAVHDVYES 207

RESULT      7
EB2928
Hypothetical protein UUI156 [imported] - Ureaplasma urealyticum
C:Species: Ureaplasma urealyticum
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: EB2928
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
Submitted to GenBank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mirror image
A:Reference number: AB2870
A:Accession: EB2928
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-153 <GLA>
A:Cross-references: GB:AE002115; GB:AF222894; NID:g6899102; PIDN:AAF30562.1; GSPDB:GN001
A:Experimental source: serovar 3; biovar 1
A:Genetics:
Gene: UUI156
Protein:
Genetic code: SGC3

Query Match          16.5%; Score 64.5; DB 2; Length 153;
Best Local Similarity 30.2%; Pred. No. 6.3;
Matches 19; Conservative 13; Mismatches 22; Indels 9; Gaps 3;

Oy 4 CKLEDMVEKTNISD--ISIPEYKELQEFIDSAAAEAMGKFQCFLNSHRT-----L 56
||| : | : | : | : : : : | : | : | : | : | : : : |
Db 49 KXIVPRMISKHLNDHLINVDYVPLTKEFIKKSVEKV--KHLYRFISVENKTKQMPEL 106
||| : | : | : | : : : : | : | : | : | : | : : : |
Oy 57 KNF 59
||
Db 107 INF 109

RESULT      8
S59558
dynamidin-like protein - Arabidopsis thaliana
N:Alternate names: GFP-binding protein, 68K
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jul-2000

```



```
Query Match      .    15.8%;   Score 62; DB 2; Length 1013;  
Best Local Similarity 28.9%; Pred. No. 2,le=02;  
Matches 22; Conservative 4; Mismatches 32; Indels 18; Gaps 1;  
  
Oy      8 EDWFEKTNDSISPEYKELOE-----FIDSDAAAEAMGFKQCFL 49  
|| :| :| | | ||  
db     1158 EDKLEEKINEDVFITEGEKKSNKIKNTQHNNNDNNDVFCINSLEYELLNKESFFL 1217  
          || :  
Oy     50 NOSHRTLKNFGLLMHT 65  
        _|_|_|||_||_  
bb     1218 NIKHGKLKYINERMHT 1233
```

Search completed: April 24, 2001, 15:14:09
Job time: 72 sec

Result No.	Score	Query Match	Length	DB ID	Description
1	159	40.6	95	11	Q9JHB9
2	147	37.5	94	11	Q9QXF3
3	139	35.5	94	11	Q9QXF2
4	76.5	19.5	280	3	Q13997
5	71	18.1	828	5	Q9VC06
6	70	17.9	1187	5	Q9V7N3
7	69.5	17.7	108	2	Q9JML7
8	68.5	17.5	487	10	Q9LVH6
9	66.5	17.0	456	10	Q9ZTV7
10	65	16.6	975	10	Q9LNB4
11	64.5	16.5	153	2	Q9PQV8
12	64.5	16.5	610	10	Q39118
13	64	16.3	1048	10	Q9SGV7
14	63.5	16.2	269	1	Q9V1S9
15	63.5	16.2	2162	5	Q9V2V4
16	63	16.1	131	2	Q3JUC2
17	62.5	15.9	496	4	Q9NTG0
18	62.5	15.9	540	5	Q9VCL3
19	62.5	15.9	694	4	Q9UPP0
20	62.5	15.9	694	4	Q9UPP0
21	62.5	15.9	694	4	Q9UPP0
22	62.5	15.9	694	4	Q9UPP0
23	62.5	15.9	694	4	Q9UPP0
24	62.5	15.9	694	4	Q9UPP0
25	62.5	15.9	694	4	Q9UPP0
26	62.5	15.9	694	4	Q9UPP0
27	62.5	15.9	694	4	Q9UPP0
28	62.5	15.9	694	4	Q9UPP0
29	62.5	15.9	694	4	Q9UPP0
30	62.5	15.9	694	4	Q9UPP0
31	62.5	15.9	694	4	Q9UPP0
32	62.5	15.9	694	4	Q9UPP0
33	62.5	15.9	694	4	Q9UPP0
34	62.5	15.9	694	4	Q9UPP0
35	62.5	15.9	694	4	Q9UPP0
36	62.5	15.9	694	4	Q9UPP0
37	62.5	15.9	694	4	Q9UPP0
38	62.5	15.9	694	4	Q9UPP0
39	62.5	15.9	694	4	Q9UPP0
40	62.5	15.9	694	4	Q9UPP0
41	62.5	15.9	694	4	Q9UPP0
42	62.5	15.9	694	4	Q9UPP0
43	62.5	15.9	694	4	Q9UPP0
44	62.5	15.9	694	4	Q9UPP0
45	62.5	15.9	694	4	Q9UPP0
46	62.5	15.9	694	4	Q9UPP0
47	62.5	15.9	694	4	Q9UPP0
48	62.5	15.9	694	4	Q9UPP0
49	62.5	15.9	694	4	Q9UPP0
50	62.5	15.9	694	4	Q9UPP0
51	62.5	15.9	694	4	Q9UPP0
52	62.5	15.9	694	4	Q9UPP0
53	62.5	15.9	694	4	Q9UPP0
54	62.5	15.9	694	4	Q9UPP0
55	62.5	15.9	694	4	Q9UPP0
56	62.5	15.9	694	4	Q9UPP0
57	62.5	15.9	694	4	Q9UPP0
58	62.5	15.9	694	4	Q9UPP0
59	62.5	15.9	694	4	Q9UPP0
60	62.5	15.9	694	4	Q9UPP0
61	62.5	15.9	694	4	Q9UPP0
62	62.5	15.9	694	4	Q9UPP0
63	62.5	15.9	694	4	Q9UPP0
64	62.5	15.9	694	4	Q9UPP0
65	62.5	15.9	694	4	Q9UPP0
66	62.5	15.9	694	4	Q9UPP0
67	62.5	15.9	694	4	Q9UPP0
68	62.5	15.9	694	4	Q9UPP0
69	62.5	15.9	694	4	Q9UPP0
70	62.5	15.9	694	4	Q9UPP0
71	62.5	15.9	694	4	Q9UPP0
72	62.5	15.9	694	4	Q9UPP0
73	62.5	15.9	694	4	Q9UPP0
74	62.5	15.9	694	4	Q9UPP0
75	62.5	15.9	694	4	

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Besley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahle C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Fostler K.Y., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RA "The genome sequence of *Drosophila melanogaster*.
 RL Science 287:2185-2195(2000).
 [2]
 SEQUENCE FROM N.A.
 Hendrick A.G., Harrington L.S.;
 "Identification of novel ARF-GTPase activating protein (GAP)-like
 proteins in the *Drosophila* genome."
 Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 EMBL; AF003742; AAF56100.1; -;
 EMBL; AF254071; AAF64529.1; -;
 HSP; P25098; 1BAK.
 DR FLYBASE; FBgn0039056; CG6742.
 DR INTERPRO; IPR001164; -;
 DR INTERPRO; IPR001547; -;
 DR INTERPRO; IPR001849; -;
 DR INTERPRO; IPR002110; -;
 DR PFAM; PF00023; ank; 2.
 DR PFAM; PF00169; PH; 1.
 DR PFAM; PF01412; ArfGAP; 1.
 DR PROSITE; PRO00405; REVINTACTING.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 DR PROSITE; PS50003; PH_DOMAIN; 1.
 DR SEQUENCE 828 AA; 93573 MW; 1CED31042512377C CRC64;

Query Match 18.1%; Score 71; DB 5; Length 828;
 Best Local Similarity 28.1%; Pred. No. 15;
 Matches 25; Conservative 14; Mismatches 24; Indels 26; Gaps 4;

QY 1 DSGCKLLEDMVKTN-SDISPEYKE-----LLQEFIDSDAAAEANGKFK 45
 Db 25 ESDIEHLEQRLEKIILCNVAVDSGKRYKNSAFAMSLWLDLQQLHFNKNAHNGKLI 84
 QY 46 QCF-----LNQSHRT-LKNFGLMM 63
 Db 85 HCFQENKFFHTILLDQASRTVLKNSVFV 113

RESULT 6
 Q9V7N3 PRELIMINARY; PRT; 1187 AA.
 ID Q9V7N3;
 AC Q9V7N3;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE CG8060 PROTEIN.
 GN CG8060.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; *Drosophila*.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY.
 RX MEDLINE=20196006; Pubmed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Besley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahle C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Fostler K.Y., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RA "The genome sequence of *Drosophila melanogaster*.
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003807; AAF58015.1; -;
 DR HSP; P42771; 1B17.
 DR FLYBASE; FBgn0034113; CG8060.
 DR INTERPRO; IPR000210; -;
 DR INTERPRO; IPR000408; -;
 DR INTERPRO; IPR002110; -;
 DR PFAM; PF00023; ank; 2.
 DR PFAM; PF00651; BTB; 1.
 DR SEQUENCE 1187 AA; 135391 MW; 7E27842BDF2C2A78 CRC64;

Query Match 17.9%; Score 70; DB 5; Length 1187;
 Best Local Similarity 27.4%; Pred. No. 28;
 Matches 23; Conservative 9; Mismatches 20; Indels 32; Gaps 3;

QY 20 SIP-----EYKELQEFIDSDAAAE-----AMGKF-----KQC 47
 Db 537 SLPRREHSFKLLHETSDCAVDHVVHVDGKFAAHKFIIVSRAPLRDLIRCYLDKDI 596
 QY 48 FLNQSHRTLNKFLMMHTVYDSIW 71
 Db 597 YLNFHDLTGKMFELILNHIYSYW 620

Copied from PCT009367009 on 27-04-200

```

SQ PFAM; PF01535; PPR; 24.
SQ SEQUENCE 1048 AA; 119167 MW; 4FB6DF1F7F56481 CRC64;

Query Match      16.3%; Score 64; DB 10; Length 1048;
Best Local Similarity 23.8%; Pred. No. 1.1e+02;
Matches 20; Conservative 19; Mismatches 11; Indels 34; Gaps 5;

QY 5 KLEDMVEKTSINDSI-----SIPE-----YKELQEFIDSDAAAEAMGKFKQ 46
232 RMLRDMKRSINPDVVTFTALIDVFVKGNLDEAQLKEMIQSSVDPN-----280
47 CFLNSHRTLNKFLMMH-TVYDS 69
281 ---NTYNSIIN-GLCMHGRLYDA 300
PAB0223.
Pyrococcus abyssi
Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
NCBI_TaxID=29292;
[1]
SEQUENCE FROM N.A.
STRAIN=ORSAY;
Heilig R.;
"Pyrococcus abyssi genome sequence: insights into archaeal chromosome
structure and evolution.";
Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
EMBL; AJ248284; CAB49270.1;
INTERPRO; IPR000217;
PROSITE; PS00228; TUBULIN_B_AUTOREG; 1.
Hypothetical protein.
SEQUENCE 269 AA; 30975 MW; DASCF9D0A5D38BDE CRC64;

Query Match      16.2%; Score 63.5; DB 1; Length 269;
Best Local Similarity 27.5%; Pred. No. 27;
Matches 22; Conservative 11; Mismatches 30; Indels 17; Gaps 4;

7 LEDMVEKTSINDSIPEYKELL-----QEFIDSDAAAEAMGKFKQCFLNQS 52
191 LEDMIE-VVESFSSPTTLTKPDENAVVRKMYENPKFVE-DVAREILMKAREKFPGR 248
53 H-RTLNFLGLMMHTVYDSIW 71
249 HVRVISNESIHKHVDIAEAW 268

RESULT 15
Q9VVA4 PRELIMINARY; PRT; 2162 AA.
AC Q9VVA4 Q9VVA3;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TRENBLrel. 14, Last annotation update)
CG9674 PROTEIN.
DE
GN Drosophila melanogaster (Fruit fly).
OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
```

```

RX MEDLINE-20196006; PubMed-10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.M., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -!- ALTERNATIVE PRODUCTS: TWO ISOFORMS; LONG ISOFORM (SHOWN HERE) AND
CC SHORT ISOFORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
DR EMBL; AE003526; RAF49409.1; -
DR EMBL; AE003526; RAF49410.1; -
DR FLXBASE; FBGN0036663; CG9674.
DR INTERPRO; IPR000103; -
DR INTERPRO; IPR000577; -
DR INTERPRO; IPR000759; -
DR INTERPRO; IPR001327; -
DR INTERPRO; IPR002489; -
DR INTERPRO; IPR002932; -
DR PFAM; PF01493; DUF14; 1.
DR PFAM; PF01645; Glu_synthase; 1.
DR PRINTS; PR00368; FADPNR.
DR PRINTS; PR00419; ADXRTASE.
DR PRINTS; PR00469; PNDRDTASEII.
DR PROSITE; PS00445; FGGY_KINASES.2; 1.
KW Alternative splicing; Hypothetical protein.
FT VARSPPLIC 1 1247 MISSING (IN SHORT ISOFORM);
SQ SEQUENCE 2162 AA; 237659 MW; E254C2EFCB0C240F CRC64;
```

Query Match 16.2%; Score 63.5; DB 5; Length 2162;
Best Local Similarity 38.9%; Pred. No. 2.8e+02;
Matches 21; Conservative 8; Mismatches 14; Indels 11; Gaps 3;

QY 5 KLEDMVEKTSINDSIPEYKELLQEFIDSDAAAEAMGKFKQCFLNQSHRTLNK 58
DB 1579 ELLADFIKT-GSKVA---KELL-----DNWAEAGKFKVFPFYEQALKD 1621

OM of: US-09-367-009-3 to: GenEmbl.* out_format : pfs
Date: Apr 28, 2001 4:44 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framed_p2n.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool/US09367009/runat_24042001_151806_19020/app_query.fasta_1.131
-DB=GenEmbl -OFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-ALIGN=15 -DOBLE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09367009_SCGN1_1.402 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-367-009-3
Query length: 74
Database: GenEmbl.*
Database sequences: 1283235
Database length: -1216004940
Search time (sec): 1355.860000

Score_list:

Sequence	Strd	Orig	ZScore	Escore	Len	Documentation
gb.pr8:HSA224173	+	390.00	900.82	5.8e-42	496	AJ224173 Homo sapiens mRNA for
gb.pat1:AX013093	+	390.00	900.43	6.1e-42	517	AX013093 Sequence 33 from Paten
gb.pat4:AF071219	+	390.00	900.43	6.1e-42	517	AF071219 Homo sapiens mammaplo
gb.htg17:AC074200	+	324.50	692.80	2.3e-30	17281	AC074200 Homo sapiens chromo
gb.htg24:AP002793	-	324.50	692.35	2.4e-30	180676	AP002793 Homo sapiens chromo
gb.htg1:AC003023	+	255.00	535.26	1.3e-21	104768	AC003023 Homo sapiens chromo
gb.om:AF308618	+	218.00	499.15	1.4e-19	506	AF308618 Oryctolagus cuniculus
gb.pat1:AR080288	+	208.00	477.96	2.1e-18	403	AR080288 Sequence 5 from patent
gb.pat1:AR095406	+	208.00	477.96	2.1e-18	403	AR095406 Sequence 5 from patent
gb.pat2:165738	+	208.00	477.96	2.1e-18	403	165738 Sequence 5 from patent
gb.pat1:AR080285	+	208.00	475.87	2.7e-18	503	AR080285 Sequence 1 from patent
gb.pat1:AR095403	+	208.00	475.87	2.7e-18	503	AR095403 Sequence 1 from patent
gb.pat2:165735	+	208.00	475.87	2.7e-18	503	165735 Sequence 1 from patent
gb.p9:HSU3147	+	208.00	475.87	2.7e-18	503	U3147 Human mammaploblin mRNA,
gb.pat4:AF015224	-	180.00	390.50	1.6e-13	4206	AF015224 Homo sapiens mammaplo
gb.htg1:AC004127	-	180.00	355.62	1.4e-11	170215	AC004127 Homo sapiens chromo
gb.om:AF308617	+	175.00	399.97	4.6e-14	17281	AC074200 Homo sapiens chromo
gb.om:AF308619	+	169.00	385.84	2.8e-13	452	AF308619 Oryctolagus cuniculus
gb.om:AF308620	+	164.00	373.98	1.3e-12	461	AF308620 Oryctolagus cuniculus
gb.ro2:RNPS07	+	163.00	370.72	2.0e-12	509	V01263 Rat mRNA encoding prosta
gb.ro1:RATPBC3	+	157.00	357.73	1.0e-11	457	J00772 rat prostatein subunit c
gb.ro1:RATPBC3	+	156.50	330.74	3.3e-10	7072	M71245 Rat prostatein C3 subun
gb.ro1:RATPBC3	+	152.00	353.47	1.8e-11	208	V01258 Rattus norvegicus DNA (a
gb.ro2:RNPS05	+	148.00	344.14	6.0e-11	208	V01261 Rattus norvegicus partia
gb.ro1:MAU252138	+	147.00	333.98	2.2e-10	477	AJ252138 Mesocricetus auratus m
gb.ro1:MAU252139	+	139.00	315.61	2.3e-09	462	AJ252139 Mesocricetus auratus m
gb.htg25:LMFLCHR34_0	+	78.50	122.81	126.61	110000	AL499623 Leishmania major ch
gb.pat3:SPAC27E2	+	76.50	129.05	56.92	34595	J298978 S.pombe chromosome I c
gb.htg18:AC084031	+	72.50	122.29	135.44	26315	AC084031 Homo sapiens chromo
gb.htg14:AC068155	+	72.50	103.95	1.4e+03	184201	AC068155 Homo sapiens chromo
gb.in2:AF254071	+	71.00	141.02	12.26	2487	AF254071 Drosophila melanogast
gb.in6:ALJ359715	+	70.50	103.13	1.6e+03	122379	ALJ359715 Human DNA sequence
gb.in4:AC012267	+	70.50	100.51	2.2e+03	161595	AC012267 Homo sapiens clone
gb.htg12:AC0207053	+	70.50	100.02	2.4e+03	170279	AC0207053 Homo sapiens chromo
gb.htg18:AC080088	+	70.50	99.56	2.5e+03	178818	AC080088 Homo sapiens chromo
gb.bas3:RPTFSY	+	70.00	130.60	46.64	5864	Y11784 R.prowazekii ftsY, pola
gb.in1:AC004318	+	70.00	105.27	1.2e+03	86188	AC004318 Drosophila melanogast
gb.htg2:AC009356	+	70.00	104.14	1.4e+03	97220	AC009356 Drosophila melanogast
gb.htg6:AC017201	+	70.00	104.12	1.4e+03	97346	AC017201 Drosophila melanogast
gb.htg1:AC008230	-	70.00	103.10	1.6e+03	108469	AC008230 Drosophila melanogast

gb.htg99:AC024034 + 70.00 100.31 2.3e+03 145961 AC024034 Homo sapiens chr
gb.htg20:AL159154 + 70.00 98.69 2.8e+03 173361 AL159154 Homo sapiens chr
gb.htg7:AC021123 + 70.00 97.63 3.2e+03 193802 AC021123 Homo sapiens chr
gb.bas3:RPXX04 + 70.00 95.72 4.1e+03 237523 AJ235273 Rickettsia prow

seq_name: gb.pr8:HSA224173

seq_documentation_block:
LOCUS HSA224173 496 bp mRNA PRI 25-MAR-1999
DEFINITION Homo sapiens mRNA for lipophilin C.
ACCESSION AJ224173
VERSION AJ224173.1 GI:4107232
KEYWORDS lipophilin C.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 496)
AUTHORS Zhao,C., Nguyen,T., Yusifov,T., Glasgow,B.J. and Lehrer,R.I.
TITLE Lipophilins: human peptides homologous to rat prostatein
JOURNAL Biochem. Biophys. Res. Commun. 256 (1), 147-155 (1999)
MEDLINE 99167354
REFERENCE 2 (bases 1 to 496)
AUTHORS Zhao,C.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-1998) Zhao C., UCLA Dept. of Medicine, CHS 37055,
Los Angeles, CA 90095, USA
FEATURES
Location/Qualifiers
source
1..496
/organism="Homo sapiens"
/db_xref="taxon:9606"
41..328
/codon_start=1
/product="lipophilin C"
/protein_id="CAA11865.1"
/db_xref="gi:4107233"
translation="MKLLMVLAAALLHCYADSGCKLLEDVKTINSDISIPYKE
LQEFIDSDAAEAEMGKFKQCLSHRTLANFLGLMHTVYDSIWNKSN"

BASE COUNT 141 a 119 c 102 g 134 t
ORIGIN
alignment_scores:
Quality: 390.00 Length: 74
Ratio: 5.270 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.649
alignment_block:
US-09-367-009-3 x HSA224173 ..
Align seg 1/1 to: HSA224173 from: 1 to: 496

1 ASPSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
|||||
95 GATTCGTGGCTGCAAACTCCTGGAGGAGCATGGTTGAAAGACCATCAATTC 144
|||||
17 rAspIleSerIleProGluTyrLysGluLeuLeuGluGluPheIleAsp 34
|||||
145 CGACATATCTATACCTACCAATACAAAGAGCTTCTCAAGAGTTCATAGACA 194
|||||
34 eRAspAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
|||||
195 GTGATCGCGCTGCAGAGGCTATGGGAAATTCAGCAGTGTTCCTCAAC 244
|||||
51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValty 67
|||||
245 CAGTCACATAGAACTCTGAAACAACTTGGACTGATGATGATGATGATGTA 294
|||||
67 rAspSerIleTyrCysAsnLeu 74
|||||
295 CGACAGCATTTGGTGAATATG 316

```

seq_name: gb_pat1:AX013093
seq_documentation_block:
LOCUS AX013093 517 bp DNA PAT 07-SEP-2000
DEFINITION Sequence 33 from Patent WO9954461.
ACCESSION AX013093
VERSION AX013093.1 GI:10040259
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 517)
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pilarsky,C.
TITLE Human nucleic acid sequences of endometrium tumour tissue
JOURNAL Patent: WO 9954461-A 33 28-OCT-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE)
FEATURES
Location/Qualifiers
source 1..517
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q13"
65..352
/note="member of the uteroglobin family"
/codon_start=1
/product="mammaglobin B precursor"
/protein_id="AAC79996.1"
/db_xref="GI:3288868"
/translation="MKLLVLMALALLHCYADSGCKLLEDVMKXTINSDISIPYK
LLQEFIDSDAAAEAMGKFKQCFNLQSHLTKNFGMLMHTVYDSIWCNKN"
BASE COUNT 156 a 121 c 106 g 134 t
ORIGIN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2559
2560
2561
2562
2563
```

TITLE The sequence of Homo sapiens clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 172281)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (12-JUL-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 COMMENT On Aug 11, 2000 this sequence version replaced gi:9247151.

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml

----- Project Information -----
 Center project name: H.NH0703H08

----- Summary Statistics -----
 Sequencing vector: M13; 100%
 Sequencing vector: plasmid; 0%

Chemistry: Dye-terminator; 100% of reads
 Chemistry: Dye-terminator Big Dye; 0% of reads

Assembly program: Phrap; version 0.990319
 Consensus quality: 166314 bases at least Q40
 Consensus quality: 168319 bases at least Q30
 Consensus quality: 169216 bases at least Q20

Insert size: 188000; agarose-fp
 Insert size: 171481; sum-of-contigs

Quality coverage: 4.93 in Q20 bases; agarose-fp
 Quality coverage: 5.45 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 9 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

 * 1 2192: contig of 2192 bp in length
 * 2193 2292: gap of unknown length
 * 2293 4860: contig of 2568 bp in length
 * 4861 4960: gap of unknown length
 * 4961 18883: contig of 13923 bp in length
 * 18884 18983: gap of unknown length
 * 18984 33466: contig of 14383 bp in length
 * 33467 33466: gap of unknown length
 * 33467 52320: contig of 18753 bp in length
 * 52320 52319: gap of unknown length
 * 52320 69757: contig of 17438 bp in length
 * 69758 69857: gap of unknown length
 * 69858 94565: contig of 24708 bp in length
 * 94566 94565: gap of unknown length
 * 94566 120988: contig of 26323 bp in length
 * 120989 121088: gap of unknown length
 * 121089 172281: contig of 51193 bp in length.

----- Location/Qualifiers -----
 1..172281
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /clone="RP11-703H8"
 BASE COUNT 49475 a 37393 c 36287 g 48325 t 801 others
 ORIGIN

FEATURES

 alignment_scores:
 Quality: 324.50 Length: 71
 Ratio: 4.992 Gaps: 1
 Percent Similarity: 91.549 Percent Identity: 91.549

----- alignment_block -----
 US-09-367-009-3 x AC074200

Align seg 1/1 to: AC074200 from: 1 to: 172281

1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
 |||||
 49571 GATTCGGCTGCAAACTCTCGAGGACATGGTTGAAAAGACCATCAATTC 49620

17 rAspIleSerIleProGluTyrIleLysGluLeuLeuGluPheIleAspS 34
 |||||
 49621 CGACATATCTATACCTGTAATACAAAGAGCTTCTCAAGAGTTTCATACAGA 49670

34 eRAspAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
 |||||
 49671 GTGATCCGCTGCAGAGGCTATGGGAAATTCAGACGATGTTTCCCTCAAC 49720

51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTy 67
 |||||
 49721 CAGTCACATAGAACTCTGAAACACTTTGGACTGATGATG..... 49759

67 rAspSerIleTyr 71
 |||||
 49760GTAATTGG 49768

seq_name: gb_hcg24:AP002793

seq_documentation_block:

LOCUS AP002793 180676 bp DNA HTG 25-OCT-2000
 DEFINITION Homo sapiens chromosome 11 clone RP11-691L4 map 11q, WORKING DRAFT
 SEQUENCE, 14 unordered pieces.

ACCESSION AP002793

VERSION AP002793.3 GI:11022730

KEYWORDS HTG; HTGS-PHASE1; HTGS-DRAFT.

SOURCE Homo sapiens DNA, clone: RP11-691L4.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 180676)

AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE Homo sapiens 180,676 genomic DNA of 11q

JOURNAL Published Only in DataBase (2000) In press

REFERENCE 2 (bases 1 to 180676)

AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE Direct Submission

JOURNAL Submitted (11-JUL-2000) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suehiro-chou, Tsukumi-Ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail: hattori@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/
 Tel:81-45-503-9111, Fax:81-45-503-9170)

COMMENT On Oct 26, 2000 this sequence version replaced gi:10801120.
 ----- Genome Center
 Center: RIKEN Genomic Sciences Center(GSC)
 Center code: RIKEN
 Web site: http://hgp.gsc.riken.go.jp/
 Contact: hattori@gsc.riken.go.jp
 ----- Project Information -----
 Center project name: Humdraft11
 Center clone name: RP11-691L4
 ----- Summary Statistics -----
 Sequencing vector: PCR products; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 175391 bases at least Q40
 Consensus quality: 177563 bases at least Q30
 Consensus quality: 178750 bases at least Q20
 Insert size: 179376; sum-of-contigs
 Quality coverage: 8.29x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
 14 contigs. The true order of the pieces is not known and their
 order in this sequence record is arbitrary. Gaps between the
 contigs are represented as runs N, but the exact sizes of the gaps
 are unknown. This record will be updated with the finished sequence

as soon as it is available and the accession number will be preserved

```

1 56642 contig of 56642 bp in length
56743 92253 contig of 35511 bp in length
92354 116265 contig of 23912 bp in length
118366 128467 contig of 12001 bp in length
128467 128467 contig of 12001 bp in length
138315 147335 contig of 8920 bp in length
147335 155435 contig of 8101 bp in length
155435 162703 contig of 7168 bp in length
162703 167910 contig of 5107 bp in length
167910 171323 contig of 3313 bp in length
171323 174108 contig of 2685 bp in length
174108 176988 contig of 2780 bp in length
176988 177089 contig of 1901 bp in length
177089 178989 contig of 1587 bp in length
178989 180676 contig of 1587 bp in length.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 56642: contig of 56642 bp in length
56643 56742: gap of 100 bp
56743 92253: contig of 35511 bp in length
92254 92353: gap of 100 bp
92354 116265: contig of 23912 bp in length
116266 116365: gap of 100 bp
116366 128366: contig of 12001 bp in length
128367 128466: gap of 100 bp
128467 138214: contig of 9748 bp in length
138215 138314: gap of 100 bp
138315 147234: contig of 8920 bp in length
147235 147334: gap of 100 bp
147335 155435: contig of 8101 bp in length
155436 155535: gap of 100 bp
155536 162703: contig of 7168 bp in length
162704 162803: gap of 100 bp
162804 167910: contig of 5107 bp in length
167911 168010: gap of 100 bp
168011 171323: contig of 3313 bp in length
171324 171423: gap of 100 bp
171424 174108: contig of 2685 bp in length
174109 174208: gap of 100 bp
174209 176988: contig of 2780 bp in length
176989 177089: gap of 100 bp
177089 178989: contig of 1901 bp in length
178990 179089: gap of 100 bp
179090 180676: contig of 1587 bp in length.

```

FEATURES

Source

```

1. .180676
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-691L4"

```

```

misc_feature
1. .56642
/note="assembly_fragment"
56743. .92253
/note="assembly_fragment"
92354. .116265
/note="assembly_fragment"
116366. .128366
/note="assembly_fragment"
128467. .138214
/note="assembly_fragment"
138315. .147234
/note="assembly_fragment"
147335. .155435
/note="assembly_fragment"
155536. .162703

```

```

misc_feature
162804. .167910
/note="assembly_fragment clone_end:SP6 vector_side:left"
misc_feature
168011. .171323
/note="assembly_fragment clone_end:T7 vector_side:left"
misc_feature
171424. .174108
/note="assembly_fragment"
misc_feature
174209. .176988
/note="assembly_fragment"
misc_feature
177089. .178989
/note="assembly_fragment"
misc_feature
179090. .180676
/note="assembly_fragment"
BASE COUNT 49650 a 39805 c 39872 g 50049 t 1300 others
ORIGIN

```

```

alignment_scores:
Quality: 324.50 Length: 71
Ratio: 4.992 Gaps: 1
Percent Similarity: 91.549 Percent Identity: 91.549

```

```

alignment_block:
US-09-367-009-3 x AP002793/rev ..

```

```

Align seg 1/1 to reverse of: AP002793 from: 1 to: 180676
1 ASPSerGlyCysLysLeuLeuGluuAspMetValGluLysThrIleAsnSe 17
|||||
92220 GATTCGGCTGCAAACTCCCTGGAGGACATGGTTGAAAGACCATCAATTC 92171
|||||
17 rASpIleSerIleProGluTyrLysGluLeuLeuGluGlnGluPheIleAsps 34
|||||
92170 CGACATATCTATACCTGAATACAAAGAGCTTCTCAAGAGTTTCATAGACA 92121
|||||
34 erASpAlaIaAlaGluuMetGlyLysPheLysGlnCysPheLeuAsn 50
|||||
92120 GTGATCGCGCTGCAGAGGCTATGGGAAATTCAGACAGTGTTCCTCAAC 92071
|||||
51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValty 67
|||||
92070 CAGTCACATAGAACTCTGAAAAAAGCTTTGGACTGATGATG..... 92032
|||||
67 rASpSerIleIleTirp 71
|||||
92031 ....GTAATTGG 92023

```

```

seq_name: gb_htgl:AC003023

```

```

seq_documentation_block:
LOCUS AC003023 104768 bp DNA HTG 21-OCT-1997
DEFINITION Homo sapiens chromosome 11 clone pDJ363p2, *** SEQUENCING IN
PROGRESS ***, 22 unordered pieces.
ACCESSION AC003023
VERSION AC003023.1 GI:2554969
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 104768)
AUTHORS Evans,G.A., Athanasiou,M., Hahner,L., Osborne-Lawrence,S.,
Franklin,T.L., Federova,N., English,C., Hinson-Cooper,S., Dunn,J.,
McFarland,J., Davie,J., Ward,T., Card,P., Patel,P., Gordon,M.,
Newton,J., Valenzuela,D., Schageman,J., Harris,J., Gotway,G.,
Syed,M., Kupfer,K., Schilling,P., Gee,V., Basit,M., Brignac,S.,
Grant,O., Bumeister,R., O'Brien,K., Garner,H.R. and Schultz,R.
Template
JOURNAL Unpublished
AUTHORS Evans,G.A., Athanasiou,M., Bradbury,P., Brignac,S., Bumeister,R.,
Davis,C., English,C., Franklin,T.L., Garner,H.R., Gordon,M.,
Gotway,G., Grant,O., Hahner,L., Harris,J., Hinson,S.,

```

Narayanaswamy, U., Newton, J., O'Brien, K., Oliver, T., Patel, P., Probst, S., Rayner, S., Schageman, J., Schilling, P., Schultz, R., Syed, M., Valenzuela, D., Ward, T., and Wilson, R.
Direct Submission
Submitted (21-OCT-1997) Genome Science and Technology Center,
University of Texas Southwestern Medical Center at Dallas, 5323
Harry Hines Blvd, Dallas, TX 75235-8591, USA
* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2212: contig of 2212 bp in length
* 2213 4510: contig of 2298 bp in length
* 4511 6531: contig of 2021 bp in length
* 6532 8749: contig of 2218 bp in length
* 8750 11232: contig of 2483 bp in length
* 11233 14278: contig of 3046 bp in length
* 14279 17433: contig of 3155 bp in length
* 17434 20944: contig of 3511 bp in length
* 20945 23332: contig of 2388 bp in length
* 23333 27127: contig of 3794 bp in length
* 27128 31030: contig of 4304 bp in length
* 31031 35667: contig of 4037 bp in length
* 35668 40459: contig of 4792 bp in length
* 40460 46343: contig of 5884 bp in length
* 46344 53576: contig of 7233 bp in length
* 53577 60895: contig of 7318 bp in length
* 60896 66779: contig of 5884 bp in length
* 66780 73895: contig of 7117 bp in length
* 73896 80224: contig of 6329 bp in length
* 80225 87195: contig of 6971 bp in length
* 87196 94843: contig of 7648 bp in length
* 94844 104768: contig of 9925 bp in length.

FEATURES
source
1..104768
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="pDJ63p2"
/chromosome="11"

BASE COUNT 27497 a 24794 c 24658 g 27771 t 48 others
ORIGIN

alignment_scores:
Quality: 255.00 Length: 63
Ratio: 4.322 Gaps: 2
Percent Similarity: 93.651 Percent Identity: 87.302

alignment_block:
US-09-367-009-3 x AC003023 ..

Align seg 1/1 to: AC003023 from: 1 to: 104768

1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
|||||
60378 GATTCGGCTGCAAACTCCGAGGACATGTTGAAAGACCATCAATC 60427

17 rAspIleSerIleProGluTyrLysGluLeuGlnGluPheIleAsp 34
|||||
60428 CGACATATCTATACCTCAATACAAAGACATCTCCAGAGTTCATGACA 60477

34 erAspAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
|||||
60478 GTGATCGCGCAGCAGAGCTATGGAATTCAGGAGTATATCTCTCAAC 60527

51 GlnSerHisArgThrLeuLysAsnPheGlyLeu 61
|||||
60528 CAGTCCCAATAAAACTCTGGAATAAATTTGGACTG 60562

seq_name: gb_om:AF308618

seq_documentation_block:
LOCUS AF308618 506 bp mRNA MAM 28-DEC-2000
DEFINITION Oryctolagus cuniculus lacrimal gland lipophilin CL2 mRNA, complete cds.

ACCESSION AF308618
VERSION AF308618.1 GI:11993599
KEYWORDS rabbit.
SOURCE

ORGANISM Oryctolagus cuniculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

REFERENCE 1 (bases 1 to 506)

AUTHORS Zhao, C., Nguyen, T.X. and Lehrer, R.I.

TITLE Rabbit Lipophilins

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 506)

AUTHORS Zhao, C., Nguyen, T.X. and Lehrer, R.I.

TITLE Direct Submission

JOURNAL Submitted (27-SEP-2000) Medicine, UCLA, 10833 Le Conte Street, Los Angeles, CA 90095, USA

FEATURES

Location/Qualifiers

source

1..506

/organism="Oryctolagus cuniculus"

/db_xref="taxon:9986"

/tissue_type="lacrimal gland"

61..342

/codon_start=1

/product="lipophilin CL2"

/protein_id="AAG42806.1"

/db_xref="GI:11993600"

/translation="MKVVMVLLLAALPLCYAGSGCVLLESWEKTIIDPSVSVBEYKA
DLQRFIDTEQTEAAVEEFKECFLSQNETLANFRVMVHTIYDSLYCAAY"

BASE COUNT 135 a 123 c 116 g 132 t

ORIGIN

alignment_scores:

Quality: 218.00 Length: 71

Ratio: 3.633 Gaps: 0

Percent Similarity: 84.507 Percent Identity: 54.930

alignment_block:

US-09-367-009-3 x AF308618 ..

Align seg 1/1 to: AF308618 from: 1 to: 506

2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSer 18
|||||

118 TCTGTTGGCTTCTTCTGGAGAGCGTCTGGAAGACCATCATC 167

18 pIleSerIleProGluTyrLysGluLeuGlnGluPheIleAspSer 35
|||||

168 GGTTCGTAGAGGAATACAAACAGACATCTCAGAGGTTCTACGAC 217

35 spAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 51
|||||

218 AGCAACCGAAGCAGCTGTAGAGGAGTTCAGAGGTCCTCCTCAGCC 267

52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyr 68
|||||

268 AGCAATGAGACTCTGGCAACTTCCGAGTCTGTCATGTCATGATATGA 317

68 pSerIleTyrCys 72
|||||

318 CAGCCTTTACTGT 330

seq_name: gb_pat1:AR080288

seq_documentation_block:

LOCUS AR080288 403 bp DNA PAT 31-AUG-2000

DEFINITION Sequence 5 from patent US 5968754.

Percent Similarity: 84.932 Percent Identity: 53.425

alignment_block:
US-09-367-009-3 x AR095406 ..

Align seg 1/1 to: AR095406 from: 1 to: 403

2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
118 TCTGGCTGCCCTTATTGGAGAAATGTGATTTCCAGAGCAATCAATCCACA 167
18 pIleSerIleProGluTyrLysGluLeuGlnGluPheIleAspSerA 35
168 AGTGCTAAGACTGAATACAAAGAACTTCTTCAGAGTTCATAGACGACA 217
35 spAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
218 ATGCCACTACAAATGCCATAGATGAATTCAGAGAAATGTTTCTTAACCAA 267
52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
268 ACGGATGAACCTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATGCA 317
68 pSerIleTrpCysAsnLeu 74
318 CAGCAGCTCTTGTGATTTA 336

seq_name: gb_pat2:165738

seq_documentation_block: 403 bp DNA PAT 07-OCT-1997
LOCUS 165738
DEFINITION Sequence 5 from patent US 5668267.
ACCESSION 165738
VERSION 165738.1 GI:2482308
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 403)
AUTHORS Watson,M.A. and Fleming,T.P.
TITLE Polynucleotides encoding mammaglobin, a mammary-specific breast cancer protein
JOURNAL Patent: US 5668267-A 5 16-SEP-1997;
FEATURES
source Location/Qualifiers
BASE COUNT 115 a 97 c 86 g 105 t
ORIGIN

alignment_scores:
Quality: 208.00 Length: 73
Ratio: 3.355 Gaps: 0
Percent Similarity: 84.932 Percent Identity: 53.425

alignment_block:
US-09-367-009-3 x I65738 ..

Align seg 1/1 to: I65738 from: 1 to: 403

2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
118 TCTGGCTGCCCTTATTGGAGAAATGTGATTTCCAGAGCAATCAATCCACA 167
18 pIleSerIleProGluTyrLysGluLeuGlnGluPheIleAspSerA 35
168 AGTGCTAAGACTGAATACAAAGAACTTCTTCAGAGTTCATAGACGACA 217
35 spAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
218 ATGCCACTACAAATGCCATAGATGAATTCAGAGAAATGTTTCTTAACCAA 267
52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
268 ACGGATGAACCTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATGCA 317
68 pSerIleTrpCysAsnLeu 74
318 CAGCAGCTCTTGTGATTTA 336

seq_name: gb_pat1:AR095406

seq_documentation_block: 403 bp DNA PAT 08-SEP-2000
LOCUS AR095406
DEFINITION Sequence 5 from patent US 6004756.
ACCESSION AR095406
VERSION AR095406.1 GI:10023252
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 403)
AUTHORS Watson,M.A. and Fleming,T.P.
TITLE Method for detecting the presence of breast cancer by detecting an increase in mammaglobin mRNA expression
JOURNAL Patent: US 6004756-A 5 21-DEC-1999;
FEATURES
source Location/Qualifiers
BASE COUNT 115 a 97 c 86 g 105 t
ORIGIN

alignment_scores:
Quality: 208.00 Length: 73
Ratio: 3.355 Gaps: 0

... ..
268 ACGGATGAACCTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATATGA 317
68 pSerIleTrpCysAsnLeu 74
318 CAGCAGCTCTTTGTGATTGA 336

seq_name: gb_pat1:AR080285

seq_documentation_block:
LOCUS AR080285 503 bp DNA 31-AUG-2000
DEFINITION Sequence 1 from patent US 5968754.
ACCESSION AR080285
VERSION AR080285.1 GI:10007020
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 503)
AUTHORS Watson,M.A. and Fleming,T.P.
TITLE Mammaglobin, a mammary-specific breast cancer protein
JOURNAL Patent: US 5968754-A 1 19-OCT-1999;
FEATURES Location/Qualifiers
source 1..503
BASE COUNT 146 a 118 c 97 g 142 t
ORIGIN

alignment_scores:
Quality: 208.00 Length: 73
Ratio: 3.355 Gaps: 0
Percent Similarity: 84.932 Percent Identity: 53.425
alignment_block:
US-09-367-009-3 x AR080285 ..
Align seg 1/1 to: AR080285 from: 1 to: 503

2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
118 TCTGGTGCCTTATGAGAAATGTTTCCAGACAAATCAATCCACA 167
18 pIleSerIleProGluTyrLysGluLeuGluPheIleAspSera 35
168 AGTGCTAAGACTGAATCAAGAAAGTCTTCAAGAGTTTCATAGAGGACA 217
35 spAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
218 ATGCCACTCAAAATGCCATAGATGAATTCAGAGAAATGTTTCTTAACCAA 267

52 SerHisArgThrLeuLysAsnPheGlyLeuMetHisThrValTyrAs 68
268 ACGGATGAACCTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATATGA 317
68 pSerIleTrpCysAsnLeu 74
318 CAGCAGCTCTTTGTGATTGA 336

seq_name: gb_pat1:AR095403

seq_documentation_block:
LOCUS AR095403 503 bp DNA 08-SEP-2000
DEFINITION Sequence 1 from patent US 6004756.
ACCESSION AR095403
VERSION AR095403.1 GI:10023248
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 503)
AUTHORS Watson,M.A. and Fleming,T.P.
TITLE Method for detecting the presence of breast cancer by detecting an

2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
118 TCTGGTGCCTTATGAGAAATGTTTCCAGACAAATCAATCCACA 167
18 pIleSerIleProGluTyrLysGluLeuGluPheIleAspSera 35
168 AGTGCTAAGACTGAATCAAGAAAGTCTTCAAGAGTTTCATAGAGGACA 217
35 spAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
218 ATGCCACTCAAAATGCCATAGATGAATTCAGAGAAATGTTTCTTAACCAA 267
52 SerHisArgThrLeuLysAsnPheGlyLeuMetHisThrValTyrAs 68
268 ACGGATGAACCTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATATGA 317
68 pSerIleTrpCysAsnLeu 74
318 CAGCAGCTCTTTGTGATTGA 336

seq_name: gb_pat1:AR095403

seq_documentation_block:
LOCUS AR095403 503 bp DNA 08-SEP-2000
DEFINITION Sequence 1 from patent US 6004756.
ACCESSION AR095403
VERSION AR095403.1 GI:10023248
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 503)
AUTHORS Watson,M.A. and Fleming,T.P.
TITLE Method for detecting the presence of breast cancer by detecting an

2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
118 TCTGGTGCCTTATGAGAAATGTTTCCAGACAAATCAATCCACA 167
18 pIleSerIleProGluTyrLysGluLeuGluPheIleAspSera 35
168 AGTGCTAAGACTGAATCAAGAAAGTCTTCAAGAGTTTCATAGAGGACA 217
35 spAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
218 ATGCCACTCAAAATGCCATAGATGAATTCAGAGAAATGTTTCTTAACCAA 267
52 SerHisArgThrLeuLysAsnPheGlyLeuMetHisThrValTyrAs 68
268 ACGGATGAACCTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATATGA 317
68 pSerIleTrpCysAsnLeu 74
318 CAGCAGCTCTTTGTGATTGA 336

increase in mammaglobin mRNA expression
Patent: US 6004756-A 1 21-DEC-1999;
FEATURES Location/Qualifiers
source 1..503
BASE COUNT 146 a 118 c 97 g 142 t
ORIGIN

alignment_scores:
Quality: 208.00 Length: 73
Ratio: 3.355 Gaps: 0
Percent Similarity: 84.932 Percent Identity: 53.425
alignment_block:
US-09-367-009-3 x AR095403 ..
Align seg 1/1 to: AR095403 from: 1 to: 503

2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
118 TCTGGTGCCTTATGAGAAATGTTTCCAGACAAATCAATCCACA 167
18 pIleSerIleProGluTyrLysGluLeuGluPheIleAspSera 35
168 AGTGCTAAGACTGAATCAAGAAAGTCTTCAAGAGTTTCATAGAGGACA 217
35 spAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
218 ATGCCACTCAAAATGCCATAGATGAATTCAGAGAAATGTTTCTTAACCAA 267
52 SerHisArgThrLeuLysAsnPheGlyLeuMetHisThrValTyrAs 68
268 ACGGATGAACCTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATATGA 317
68 pSerIleTrpCysAsnLeu 74
318 CAGCAGCTCTTTGTGATTGA 336

seq_name: gb_pat2:I65735
seq_documentation_block:
LOCUS I65735 503 bp DNA PAT 07-OCT-1997
DEFINITION Sequence 1 from patent US 5668267.
ACCESSION I65735
VERSION I65735.1 GI:2482305
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 503)
AUTHORS Watson,M.A. and Fleming,T.P.
TITLE Polynucleotides encoding mammaglobin, a mammary-specific breast cancer protein
JOURNAL Patent: US 5668267-A 1 16-SEP-1997;
FEATURES Location/Qualifiers
source 1..503
BASE COUNT 146 a 118 c 97 g 142 t
ORIGIN

alignment_scores:
Quality: 208.00 Length: 73
Ratio: 3.355 Gaps: 0
Percent Similarity: 84.932 Percent Identity: 53.425
alignment_block:
US-09-367-009-3 x I65735 ..
Align seg 1/1 to: I65735 from: 1 to: 503

2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
118 TCTGGTGCCTTATGAGAAATGTTTCCAGACAAATCAATCCACA 167
18 pIleSerIleProGluTyrLysGluLeuGluPheIleAspSera 35
168 AGTGCTAAGACTGAATCAAGAAAGTCTTCAAGAGTTTCATAGAGGACA 217
35 spAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
218 ATGCCACTCAAAATGCCATAGATGAATTCAGAGAAATGTTTCTTAACCAA 267
52 SerHisArgThrLeuLysAsnPheGlyLeuMetHisThrValTyrAs 68
268 ACGGATGAACCTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATATGA 317
68 pSerIleTrpCysAsnLeu 74
318 CAGCAGCTCTTTGTGATTGA 336

seq_name: gb_pat1:AR095403
seq_documentation_block:
LOCUS AR095403 503 bp DNA PAT 08-SEP-2000
DEFINITION Sequence 1 from patent US 6004756.
ACCESSION AR095403
VERSION AR095403.1 GI:10023248
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 503)
AUTHORS Watson,M.A. and Fleming,T.P.
TITLE Method for detecting the presence of breast cancer by detecting an

2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
118 TCTGGTGCCTTATGAGAAATGTTTCCAGACAAATCAATCCACA 167
18 pIleSerIleProGluTyrLysGluLeuGluPheIleAspSera 35
168 AGTGCTAAGACTGAATCAAGAAAGTCTTCAAGAGTTTCATAGAGGACA 217
35 spAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
218 ATGCCACTCAAAATGCCATAGATGAATTCAGAGAAATGTTTCTTAACCAA 267
52 SerHisArgThrLeuLysAsnPheGlyLeuMetHisThrValTyrAs 68
268 ACGGATGAACCTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATATGA 317
68 pSerIleTrpCysAsnLeu 74
318 CAGCAGCTCTTTGTGATTGA 336

seq_name: gb_pat1:AR095403
seq_documentation_block:
LOCUS AR095403 503 bp DNA PAT 08-SEP-2000
DEFINITION Sequence 1 from patent US 6004756.
ACCESSION AR095403
VERSION AR095403.1 GI:10023248
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 503)
AUTHORS Watson,M.A. and Fleming,T.P.
TITLE Method for detecting the presence of breast cancer by detecting an

2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
118 TCTGGTGCCTTATGAGAAATGTTTCCAGACAAATCAATCCACA 167
18 pIleSerIleProGluTyrLysGluLeuGluPheIleAspSera 35
168 AGTGCTAAGACTGAATCAAGAAAGTCTTCAAGAGTTTCATAGAGGACA 217
35 spAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
218 ATGCCACTCAAAATGCCATAGATGAATTCAGAGAAATGTTTCTTAACCAA 267
52 SerHisArgThrLeuLysAsnPheGlyLeuMetHisThrValTyrAs 68
268 ACGGATGAACCTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATATGA 317
68 pSerIleTrpCysAsnLeu 74
318 CAGCAGCTCTTTGTGATTGA 336

2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
118 TCTGGTGCCTTATGAGAAATGTTTCCAGACAAATCAATCCACA 167
18 pIleSerIleProGluTyrLysGluLeuGluPheIleAspSera 35
168 AGTGCTAAGACTGAATCAAGAAAGTCTTCAAGAGTTTCATAGAGGACA 217
35 spAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
218 ATGCCACTCAAAATGCCATAGATGAATTCAGAGAAATGTTTCTTAACCAA 267
52 SerHisArgThrLeuLysAsnPheGlyLeuMetHisThrValTyrAs 68
268 ACGGATGAACCTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATATGA 317
68 pSerIleTrpCysAsnLeu 74
318 CAGCAGCTCTTTGTGATTGA 336


```

118 TCTGCGTCCCTTTATTGGAGAATGTGATTTCCAAGCAATCAATCCACA 167
      |||||    |||||||    :|||||    :|||||    :|||
18 pIeSerlleProGLutryLYsgluLeuleuGlnGlupheIIeaSpSera 35
      :|||||    :|||||    :|||||    :|||||    :|||
168 AGTGTC TAAGACTGTAATACAAAGAACTCTTCAAAGAGTTCATAGAGCACA 217
      :|||||    :|||||    :|||||    :|||||    :|||
35 sPaAlaalaalaGluaImetGlyLYspheLYsgInCysPheLeuAsnGln 51
      :|||||    :|||||    :|||||    :|||||    :|||
218 ATGCCACTACAATGCCATAGATGAATTGAAGGAATGTTTTCTTAACCACA 267
      :|||||    :|||||    :|||||    :|||||    :|||
52 SerHisArGThrLeuLYsasnPheGlyLeuMetMetHisThrvAltyras 68
      :|||    :|||    :|||    :|||    :|||
268 ACGGATGNAACCTGAGCAACTGTTGAGGTGTTTATGCAATTAATATATGA 317
      :|||    :|||    :|||    :|||    :|||
68 pSerlleleTrpCysAsnLeu 74
      |||||
318 CAGCAGCTCTTTGTGATTTA 336

```

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

```

1..303
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="breast"

61..342
/note="similar to uteroglobin: SwissProt Accession Number
P02779 and rat prostatic steroid binding protein subunit
c3: SwissProt Accession Number P02780"
/codon_start=1
/product="mammaglobin"
/protein_id="AAC50375.1"
/db_xref="gi:1199596"
/translation="MKLLMVLMAALSOHCYAGSGCPLENIKSTINPQVSKTEYKE
LLQEFIDNATTNAIDELKECFINQTDFTLSNVFVFMQLIYDSSLCDLF"

146 a 118 c 97 g 142 t
BASE COUNT
ORIGIN

alignment_scores:
    Quality: 208.00    Length: 73
    Ratio: 3.355      Gaps: 0
    Percent Similarity: 84.932    Percent Identity: 53.425

alignment_block:
    US-09-367-009-3 x HSU33147 ..
    Align seg 1/1 to: HSU33147 from: 1 to: 503

```

```
alignment_block:  x HSU333147
```

```

Percent Similarity: 83.824    Percent Identity: 51.471

alignment_block:
US-09-367-009-3 x AF015224    ..

Align seg 1/1 to: AF015224 from: 1 to: 4206

      2  SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
      1715  TCTGGCTGCCCTTTATTGGAGAAATGTGATTTCCAAAGACAATCAATCCACA 1764
      18  pileSerIleProGluTyrLysGluLeuLeuGluGlnGluPheIleAspSerA 35
      1765  AGTCCTCAAGACTCAATACAAAGAACTTCTTCAAGAGTTCATAGACGACA 1814
      35  spAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
      1815  ATGCCACTACAAGTGCCCATGATGAATTGAAGGAATGTTTCTTAAACCAA 1864
      52  SerHisArgThrLeuLysAsnPhcelyLeuMetMetHisThrValTyrAs 68
      1865  ACGATGAACACTCTGAGCAATGTTTCAGGTGTTTATGTAATTTTCATTTTC 1914

      68  pSer 69
      1915  TTCC 1918

```

OM of: US-09-367-009-3 to: N_Geneseq_0401:* out_format : pfs
Date: Apr 28, 2001 4:48 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=framed_p2n.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US09367009/runat_24042001_151807_19057/app_query.fasta_1.131
-DB=N_Geneseq_0401 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-XGAPOP=6.000 -XGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blossum62
TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=pct
THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
NORM-ext -MINLEN=0 -MAXLEN=200000000
USER=US09367009_ECGN1_1_174 -NCPU=5 -ICPU=3 -LONGLOG -NO_XLPXY
WAIT -THREADS=1

Search information block:
Query: US-09-367-009-3
Query length: 74
Database: N_Geneseq_0401:*
Database sequences: 678276
Database length: 291890651
Search time (sec): 112.140000

core_list:	Strd Orig	ZScore	Escore	Len	Documentation
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:243008 +	390.00	1007.39	3.6e-48	473	
/SID52/gcgdata/geneseq/geneseq/NA1997.DAT:794832 +	390.00	1007.32	3.6e-48	476	
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:A59730 +	390.00	1007.32	3.6e-48	476	
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:A09104 +	390.00	1007.32	3.6e-48	476	
/SID52/gcgdata/geneseq/geneseq/NA1999.DAT:X36138 +	390.00	1007.12	3.7e-48	485	
/SID52/gcgdata/geneseq/geneseq/NA1999.DAT:A24203 +	390.00	1006.45	4.0e-48	517	
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:A09118 +	390.00	1006.45	4.0e-48	517	
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:A12635 +	208.00	534.76	7.6e-22	279	
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:A12635 +	208.00	530.87	1.2e-21	403	
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:A12635 +	208.00	528.69	1.6e-21	495	
/SID52/gcgdata/geneseq/geneseq/NA1998.DAT:A11580 +	208.00	528.52	1.7e-21	503	
/SID52/gcgdata/geneseq/geneseq/NA1997.DAT:T50925 +	208.00	528.52	1.7e-21	503	
/SID52/gcgdata/geneseq/geneseq/NA1998.DAT:A17905 +	208.00	528.52	1.7e-21	503	
/SID52/gcgdata/geneseq/geneseq/NA1999.DAT:X26966 +	208.00	528.52	1.7e-21	503	
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:A12632 +	208.00	528.52	1.7e-21	503	
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:F17694 +	208.00	528.36	1.7e-21	511	
/SID52/gcgdata/geneseq/geneseq/NA1998.DAT:F17694 +	208.00	527.62	1.9e-21	548	
/SID52/gcgdata/geneseq/geneseq/NA2001.DAT:F17729 +	185.00	471.75	2.4e-18	356	
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:A81460 +	67.00	118.67	113.56	20844	
/SID52/gcgdata/geneseq/geneseq/NA1998.DAT:A44253 +	66.00	147.75	2.72	1038	
/SID52/gcgdata/geneseq/geneseq/NA1999.DAT:X99624 +	64.50	152.37	1.51	462	
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:A01735 +	64.50	147.19	2.93	754	
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:A04971 +	64.50	137.80	9.76	1833	
/SID52/gcgdata/geneseq/geneseq/NA1988.DAT:N80437 +	64.50	120.71	87.37	9229	
/SID52/gcgdata/geneseq/geneseq/NA1999.DAT:Z17473 +	64.00	145.67	3.56	769	
/SID52/gcgdata/geneseq/geneseq/NA1999.DAT:Z17473 +	64.00	145.67	3.56	769	
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:A70106 +	62.50	117.71	128.40	7458	
/SID52/gcgdata/geneseq/geneseq/NA1999.DAT:X32482 +	62.50	116.78	144.68	8144	
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:A70217 +	62.00	128.86	30.74	2295	
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:A70219 +	62.00	118.62	114.23	6042	
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:A81512 +	62.00	115.53	169.78	8093	
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:F21609 +	62.00	75.71	2.8e+04	349980	
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:F21609 +	62.00	75.71	2.8e+04	349980	
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:A81490 +	62.00	60.77	1.8e+05	1437668	
/SID52/gcgdata/geneseq/geneseq/NA1998.DAT:V26083 +	61.50	121.72	76.81	3982	
/SID52/gcgdata/geneseq/geneseq/NA1998.DAT:V13935 +	61.50	112.12	263.00	9870	
/SID52/gcgdata/geneseq/geneseq/NA1998.DAT:V16457 +	61.50	112.12	263.00	9870	
/SID52/gcgdata/geneseq/geneseq/NA1999.DAT:X13035 +	61.50	106.45	544.02	16870	
/SID52/gcgdata/geneseq/geneseq/NA1998.DAT:V26084 +	61.50	94.56	2.5e+03	51952	
/SID52/gcgdata/geneseq/geneseq/NA1998.DAT:V11152 +	61.00	125.26	48.75	2515	
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:A12387 +	60.50	134.37	15.15	938	

/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:C49309 + 60.50 129.30 29.04 1516
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:C41123 + 60.50 129.28 29.10 1518
/SID52/gcgdata/geneseq/geneseq/NA1998.DAT:V11858 + 60.50 128.09 33.92 1700
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:A81516 + 60.00 117.29 135.51 4169
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:F21611 + 60.00 70.45 5.4e+04 349980

seq_name: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT:243008

seq_documentation_block:
ID 243008 standard; CDNA; 473 BP.

AC 243008;
DT 01-FEB-2000 (first entry)

XX Human 5' EST isolated from a cDNA library SEQ ID NO:767.

XX Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW gene therapy; chromosome mapping; upstream regulatory sequence;
KW forensic; location; development; protein synthesis; stability;
KW regulation; identification; ss.

XX Homo sapiens.

XX WO9953051-A2.

XX 21-OCT-1999.

XX 09-APR-1999; 99WO-IB00712.

XX 09-APR-1998; 98US-0057719.

XX 28-APR-1998; 98US-0069047.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-038446/03.

XX P-PSDB; Y65394.

XX Novel secreted protein 5' expressed sequence tag sequences used in
diagnostic, forensic, gene therapy, and chromosome mapping procedures
Claim 1; Page 569; 837pp; English.

242265 to 243075 represent novel 5' expressed sequence tag (EST)
sequences, corresponding to human secreted proteins. Y64651 to Y65438
represent the EST-related proteins corresponding to 242265 to 243052.
The 5' ESTs can be used for producing secreted human gene products.
They can be used to identify and isolate 5' untranslated regions (UTRs)
and upstream regulatory regions which control the location, development
and rate, and quantity of protein synthesis, as well as stability of
mRNA. The ESTs are also useful as probes for chromosome mapping, and to
obtain full length cDNA clones. The ESTs can also be used in forensic
procedures to identify individuals, or in diagnostic procedures to
identify individuals having genetic diseases resulting from abnormal
gene expression. The products may also be used in gene therapy protocols.
The nucleic acids encoding signal peptides can be used for directing
extracellular secretion of a polypeptide or the insertion of a
polypeptide into a membrane, or importing a polypeptide into a cell.
The proteins encoded by the EST sequences may be useful in treating a
variety of human conditions. Secreted proteins have therapeutic value,
and the identification of new secreted proteins is valuable. 242249 to
242264 and Y64644 to Y64650 represent sequences used in the
exemplification of the present invention.

XX Sequence 473 BP; 130 A; 113 C; 104 G; 123 T; 3 other;

alignment_scores:
Quality: 390.00 Length: 74
Ratio: 5.270 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.649

```
alignment_block:
US-09-367-009-3 x 243008
Align seg 1/1 to: 243008 from: 1 to: 473

1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
|||||
114 GATTCTGGCTGCAAACTCTGGAGGACATGGTTGAAAGACCATCAATC 163
|||||
17 rAspIleSerIleProGluTyrLysGluLeuLeuGlnGluPheIleAsps 34
|||||
164 CGACATATCTATACCTGAATACAAAGAGCTCTTCAGAGCTTCATAGACA 213
|||||
34 erAspAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
|||||
214 GTGATGCCGCTGCAGAGGCTATGGGAAATTCAGCAGTGTTCCTCAAC 263
|||||
51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTy 67
|||||
264 CAGTCACATAGAACTCTGAAAACTTTGGACTGATGATGATACATAGTGA 313
|||||
67 rAspSerIleTrpCysAsnLeu 74
|||||
314 CGACAGCATTTGGTGTAAATG 335
|||||

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:T94832
seq_documentation_block:
T94832 standard; DNA: 476 BP.
T94832;

27-MAR-1998 (first entry)
Human endometrial specific steroid-binding factor III DNA.

Endometrial specific steroid-binding factor III; ESF III; human;
endometrium; phospholipase A2 inhibitor; polychlorinated biphenyl;
antiagregant; inflammation; asthma; rhinitis; cystic fibrosis;
airway disease; neoplasia; atopy; therapy; diagnosis; ss.

Homo sapiens.

Key Location/Qualifiers
CDS 46..333
/*tag= a
sig_peptide 46..108
/*tag= b
mat_peptide 109..330
/*tag= c

WO9734997-A1.
25-SEP-1997.
21-MAR-1996; 96WO-US03857.
21-MAR-1996; 96WO-US03857.
(HUMA-) HUMAN GENOME SCI INC.
Gentz RL, Ni J, Yu G;
WPI; 1997-480206/44.
P-PSDB; W35804.
Human endometrial specific steroid-binding factor I, II and III -
used to treat inflammation, asthma, rhinitis, cystic fibrosis,
airway disease, neoplasia, atopy etc.

Claim 11; Page 64; 92pp: English.
```

```
CC This polynucleotide, deposited as ATCC 97403, encodes human
CC endometrial specific steroid binding factor III (ESF III, see
CC W35804), a protein that inhibits phospholipase A2 activity, binds
CC to polychlorinated biphenyl compounds, reduces foreign protein
CC antigenicity, inhibits monocyte and neutrophil chemotaxis and
CC phagocytosis, inhibits platelet aggregation, regulates eicosanoid
CC levels in the human uterus and controls the growth of endometrial
CC cells. The polynucleotide was discovered in a cDNA library derived
CC from a human endometrial tumour. ESF I (see T94830) and ESF II
CC (see T94831) clones are also claimed. Human ESF III has about
CC 36% identity with rat prostatic steroid-binding protein. ESF I, II
CC and III polynucleotides can be used in the production of
CC recombinant polypeptides in host cells, and in claimed methods for
CC (a) the treatment of a patient in need of ESF I, II or III and (b)
CC for the diagnosis of a disease or a susceptibility to a disease
CC related to underexpression of ESF I, II or III. hESF I, II and III
CC may be used to treat inflammation, asthma, rhinitis, cystic
CC fibrosis, airway disease, neoplasia and atopy.
XX
SQ Sequence 476 BP; 131 A; 115 C; 105 G; 125 T; 0 other;

alignment_scores:
Quality: 390.00 Length: 74
Ratio: 5.270 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.649

alignment_block:
US-09-367-009-3 x T94832
Align seg 1/1 to: T94832 from: 1 to: 476

1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
|||||
100 GATTCTGGCTGCAAACTCTCTGGAGGACATGGTTGAAAGACCATCAATC 149
|||||
17 rAspIleSerIleProGluTyrLysGluLeuLeuGlnGluPheIleAsps 34
|||||
150 CGACATATCTATACCTGAATACAAAGAGCTCTTCAGAGCTTCATAGACA 199
|||||
34 erAspAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
|||||
200 GTGATGCCGCTGCAGAGGCTATGGGAAATTCAGCAGTGTTCCTCAAC 249
|||||
51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTy 67
|||||
250 CAGTCACATAGAACTCTGAAAACTTTGGACTGATGATGATACATAGTGA 299
|||||
67 rAspSerIleTrpCysAsnLeu 74
|||||
300 CGACAGCATTTGGTGTAAATG 321
|||||

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:A59730
seq_documentation_block:
ID A59730 standard; cDNA; 476 BP.
XX
AC A59730;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human endometrial specific steroid-binding factor III coding sequence.
XX
KW Endometrial specific steroid-binding factor; human; hESF; inflammation;
KW asthma; rhinitis; cystic fibrosis; air way disease; neoplasia; atopy;
KW eicosanoid level regulator; chemotaxis inhibitor; endometrial cancer; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 46..333
FT /*tag= a
FT /*product= "hESF III"
```

```
FT /note= "Endometrial specific steroid-binding factor III"
XX US606724-A.
XX
XX 23-MAY-2000.
XX
XX 21-MAR-1997; 97US-0821451.
XX
XX 21-MAR-1996; 96US-0014724.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Yu G, Gentz R, Ni J;
XX WPI; 2000-375600/32.
XX P-PSDB; B03769.
XX
XX Novel gene encoding human endometrial specific steroid-binding factor
XX I, II and III which is useful for treating asthma, rhinitis, cystic
XX fibrosis, airway disease and neoplasia
XX
XX Claim 1; Fig 3; 36pp; English.
XX
XX This invention relates to nucleic acid molecules encoding portions of the
XX human endometrial specific steroid-binding factors I, II, and III. Also
XX included in the invention are hESF I, II, and III polypeptide sequences.
XX The nucleotide sequence exhibit antiasthmatic, antiinflammatory,
XX antiallergic, and cytostatic properties. The polypeptides are used in
XX gene therapy to express hESF I, II and III polypeptides in vivo to treat
XX and/or prevent inflammation, asthma, rhinitis, cystic fibrosis, air way
XX disease, neoplasia and atopy. The polynucleotides are also used to
XX inhibit phospholipase A2 activity, bind polychlorinated biphenyls, reduce
XX foreign protein antigenicity, inhibit monocyte and neutrophil chemotaxis
XX and phagocytosis, inhibit platelet aggregation, regulate eicosanoid
XX levels in the human uterus and control the growth of endometrial cells.
XX The polynucleotides are also useful for detecting complementary
XX polynucleotides as a diagnostic reagent. The hESF I, II and III
XX polynucleotides are used to detect complementary polynucleotides such as
XX a diagnostic reagent. Detection of a mutated form of hESF I, II and III
XX associated with a dysfunction will provide a diagnostic tool that can
XX define diagnosis of a disease or susceptibility to a disease which
XX results from under-expression, over-expression or altered expression of
XX hESF I, II and III e.g. a susceptibility to inherited asthma and
XX endometrial cancer. They are also useful for chromosome identification.
XX The present sequence represents a hESF III coding sequence identified in
XX the invention.
XX
XX Sequence 476 BP; 131 A; 115 C; 105 G; 125 T; 0 other;
XX
XX
XX alignment_scores:
XX Quality: 390.00 Length: 74
XX Ratio: 5.270 Gaps: 0
XX Percent Similarity: 100.000 Percent Identity: 98.649
XX
XX alignment_block:
XX US-09-367-009-3 x A59730 ..
XX
XX Align seg 1/1 to: A59730 from: 1 to: 476
XX
XX 1 AspSerGlyCysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
XX
XX 100 GATTCGGCTGCAAACTCTGGAGGACATGGTTGAAAGACCACATCAATTC 149
XX
XX
XX 17 rAspSerIleProGluTyrLysGluLeuLeuGluGlnPheIleAspS 34
XX
XX 150 CGACATATCATACCTGCAATPACAAAGAGCTTCTCAAGAGTTCATGACA 199
XX
XX
XX 34 erAspAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
XX
XX 200 GTGATCCCGCTGAGAGGCTATGGGAAATTCAGCAGTGTTCCTCAAC 249
XX
XX
XX 51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMethIsthrValTy 67
```

```

XX
XX 250 CAGTCACATAGAACTCTGAAAACTTTGGACTGATGATGCTACAGTGTA 299
XX
XX 67 rAspSerIleTrpCysAsnLeu 74
XX
XX 300. CGACAGCATTTGGTGAATATG 321
XX
XX
XX seq_name: /SIDS2/cgdata/geneseq/geneseq/NA2000.DAT:A09104
XX
XX seq_documentation_block:
XX ID A09104 standard; DNA: 476 BP.
XX
XX AC A09104;
XX
XX DT 10-AUG-2000 (first entry)
XX
XX DE Human endometrial specific steroid binding factor coding sequence.
XX
XX KW Endometrial specific steroid binding factor; ESBPII; diagnosis;
XX gynaecological cancer; uterine; breast; endometrial; ovarian;
XX antibody; gene therapy; cytostatic; ss.
XX
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
XX CDS 46..333
XX FT /*tag= a
XX
XX PN WO200020044-A1.
XX
XX PD 13-APR-2000.
XX
XX PF 30-SEP-1999; 99WO-US22753.
XX
XX PR 02-OCT-1998; 98US-0102743.
XX
XX PA (DIAD-) DIADEXUS LLC.
XX
XX PI Macina RA;
XX
XX DR WPI; 2000-303649/26.
XX P-PSDB; Y92226.
XX
XX PT Diagnosing, staging and monitoring gynecological cancer comprising
XX using an elevated level of ESBPII in a patient as an indicator of
XX cancer
XX
XX PS Claim 6; Page 28; 32pp; English.
XX
XX CC The levels of human endometrial specific steroid binding factor
XX (ESBPII) can be measured and compared to control levels and used to
XX diagnose the presence of a gynaecological (uterine, breast, endometrial,
XX or ovarian) cancer in a patient. ESBPII levels can also be used to
XX diagnose metastasis, to stage or monitor gynaecological cancer.
XX
XX CC Antibodies specific for ESBPII can be used to treat gynecological
XX cancers.
XX
XX SQ Sequence 476 BP; 131 A; 115 C; 105 G; 125 T; 0 other;
XX
XX
XX alignment_scores:
XX Quality: 390.00 Length: 74
XX Ratio: 5.270 Gaps: 0
XX Percent Similarity: 100.000 Percent Identity: 98.649
XX
XX alignment_block:
XX US-09-367-009-3 x A09104 ..
XX
XX Align seg 1/1 to: A09104 from: 1 to: 476
XX
XX 1 AspSerGlyCysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
XX
XX 100 GATTCGGCTGCAAACTCTGGAGGACATGGTTGAAAGACCACATCAATTC 149
XX
XX
XX 17 rAspSerIleProGluTyrLysGluLeuLeuGluGlnPheIleAspS 34
XX
XX 150 CGACATATCATACCTGCAATPACAAAGAGCTTCTCAAGAGTTCATGACA 199
XX
XX
XX 34 erAspAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
XX
XX 200 GTGATCCCGCTGAGAGGCTATGGGAAATTCAGCAGTGTTCCTCAAC 249
XX
XX
XX 51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMethIsthrValTy 67
```


XX This invention describes novel human nucleic acid (cDNA) sequences (A),
 CC that are highly expressed in uterine tumour tissue and which have
 CC anticancer and cytostatic activity. (A) are used (i) for recombinant
 CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)
 CC are used (i) to identify agents suitable for treatment of uterine or
 CC endometrial cancer; (ii) directly for treating these forms of cancer
 CC (including expression from gene therapy vectors) and (iii) for generation
 CC of specific antibodies. (A) are identified by assembling ESTs (expressed
 CC sequence tags) from a particular tissue type before comparison of
 CC expression patterns. This allows a significantly longer fragment of the
 CC gene to be revealed, so should reduce the number of failures associated
 CC with the fact that ESTs from different libraries may represent different
 CC parts of the same known gene, distorting the estimated frequency of
 CC occurrence in a particular tissue. 241981-242121 represent EST fragments
 CC derived from a human endometrium tumour cDNA library which encode the
 CC protein sequences represented in Y5941-Y60328.

XX Sequence 517 BP; 156 A; 121 C; 106 G; 134 T; 0 other;

alignment_scores:
 Quality: 390.00 Length: 74
 Ratio: 5.270 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 98.649

alignment_block:
 US-09-367-009-3 x 242013 ..

Align seg 1/1 to: 242013 from: 1 to: 517

1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
 |||||
 95 GATTCTGGCTGCAAACTCTCGAGGACATGTTGAAAGACCATCAATTC 144

17 rAspIleSerIleProGluTyrLysGluLeuLeuGlnGluPheIleAsps 34
 |||||
 145 CGACATATCTATACCTGAATACAAAGAGCTTCTTCAAGAGTTTCATAGACA 194

34 erAspAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
 |||||
 195 GTGATGCCGCTGCGAGGCTATGGGAAATTCAGACAGTGTTCCTCAAC 244

51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTy 67
 |||||
 245 CAGTCACATAGAACTCTGAAAACCTTTGGACTGATGATGATACAGTGTA 294

67 rAspSerIleTrpCysAsnLeu 74
 |||||
 295 CGACAGCATTTGGTGAATATG 316

seq_name: /SDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:A09118

seq_documentation_block:

D A09118 standard; cDNA; 517 BP.

XX

AC A09118;

XX

BT 10-AUG-2000 (first entry)

XX

DE Clone Mamm-X cDNA, encodes mamoglobin homologue.

XX

KW Clone Mamm-X; mamoglobin; breast cancer; cytostatic; anti-HIV;
 immunosuppressive; antiallergic; antineoplastic; antiinflammatory;
 antiarthritic; antiarteriosclerotic; vasotropic; neuroprotective;
 neurotropic; dermatological; tranquilizer; vulnerary; ss.

OS Homo sapiens.

XX

FH Key Location/Qualifiers

XX CDS 65..352

FT /*tag= a

XX

PN WO200020447-A2.
 PD
 XX 13-APR-2000.
 XX PF 06-OCT-1999; 99WO-US23294.
 XX PR 06-OCT-1998; 98US-0103195.
 XX PR 05-OCT-1999; 99US-0103195.
 XX (CURA-) CURAGEN CORP.
 XX Shimketa RA;
 XX WPI; 2000-303741/26.
 XX P-PSDB; Y92237.
 XX Nucleic acids encoding polypeptides with syncollin-like, claudin-like or
 XX cytokine-like activity, useful for treating diseases including cancer,
 XX Alzheimer's and atherosclerosis
 XX Claim 28; Fig 6; 118pp; English.
 XX Clone Mamm-X encodes a polypeptide that is 100 percent identical to human
 CC Mamoglobin B precursor, a potential marker of breast cancer nodal
 CC metastasis. The sequences are useful for treatment of diseases such as
 CC cancer, immune disorders, autoimmune disease, acquired immune deficiency
 CC syndrome (AIDS), transplant rejection, allergy, infection by a
 CC pathological agent or organism, inflammatory disorders, arthritis, a
 CC haematopoietic disorder, a skin disorder, atherosclerosis, restenosis, a
 CC neurological disease, Alzheimer's disease, trauma, spinal cord injury
 CC and skeletal disorders.
 XX Sequence 517 BP; 144 A; 129 C; 105 G; 139 T; 0 other;

alignment_scores:

Quality: 390.00 Length: 74

Ratio: 5.270 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 98.649

alignment_block:

US-09-367-009-3 x A09118 ..

Align seg 1/1 to: A09118 from: 1 to: 517

1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
 |||||
 119 GATTCTGGCTGCAAACTCTCGAGGACATGTTGAAAGACCATCAATTC 168

17 rAspIleSerIleProGluTyrLysGluLeuLeuGlnGluPheIleAsps 34
 |||||
 169 CGACATATCTATACCTGAATACAAAGAGCTTCTTCAAGAGTTTCATAGACA 218

34 erAspAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
 |||||
 219 GTGATGCCGCTGCGAGGCTATGGGAAATTCAGACAGTGTTCCTCAAC 268

51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTy 67
 |||||
 269 CAGTCACATAGAACTCTGAAAACCTTTGGACTGATGATGATACAGTGTA 318

67 rAspSerIleTrpCysAsnLeu 74
 |||||
 319 CGACAGCATTTGGTGAATATG 340

seq_name: /SDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:A12642

seq_documentation_block:

ID A12642 standard; cDNA; 279 BP.

XX AC A12642;

XX

DT 25-JUL-2000 (first entry)


```

||||| |||:||||
318 CAGCAGTCTTTGTGATTTA 336

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:V41580
seq_documentation_block:
ID V41580 standard; CDNA; 495 BP.
XX AC V41580;
XX DT 12-OCT-1998 (first entry)
XX DE Nucleotide sequence of the human steroid binding protein C2.
XX KW Human steroid-binding protein C2; hSBP2; hSBP1; breast cancer; probe;
XX KW gene therapy vector; ribozyme; probe; hybridisation; amplification;
XX KW antibody; immunoassay; ds.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX CDS 46..327
XX /tag= a
XX /product= "human steroid binding protein C2"

WO9821331-A1.
XX 22-MAY-1998.
XX 07-NOV-1997; 97WO-US20674.
XX 12-NOV-1996; 96US-0747547.
XX (INCY-) INCYTE PHARM INC.
XX Akerblom IE, Goli SK, Hawkins PR, Hillman JL, Murry LE;
XX WPI: 1998-297935/26.
XX P-PSDB; W59777.
XX New human steroid binding proteins C1 and C2 - useful for, e.g.
XX diagnosis, monitoring and treating breast cancer, and for drug
XX screening
XX Claim 14; Fig 2; 70pp; English.
XX This is the nucleotide sequence of the human steroid-binding protein
XX C2 (hSBP2) used in the method of the invention for the diagnosis,
XX monitoring and treatment of breast cancer. HSBP1 and hSBP2 are useful
XX as markers for breast cancer, i.e. measuring levels of hSBP1 and hSBP2
XX used for diagnosis or monitoring the disease, to identify subjects
XX at risk and to discriminate between different forms of cancer for
XX selection of appropriate therapies. They may also be used for drug
XX screening. Nucleic acids encoding hSBP1 and hSBP2 can be used in gene
XX therapy vectors to overexpress the steroid-binding proteins, preventing
XX binding of steroids, or antisense sequences, ribozymes. Their nucleic
XX acids can also be used for the diagnosis and monitoring (by quantifying
XX expression of hSBP), as source of probes for hybridisation and
XX CC amplification of genomic or related sequences for studying regulation of
XX CC gene function and for mapping the genomic sequence. Antibodies are used
XX as diagnostic reagents in standard immunoassays for hSBP.
XX Sequence 495 BP; 148 A; 114 C; 94 G; 139 T; 0 other;

alignment_scores:
Quality: 208.00 Length: 73
Ratio: 3.355 Gaps: 0
Percent Similarity: 84.932 Percent Identity: 53.425

alignment_block:
US-09-367-009-3 x V41580

```

Align seg 1/1 to: V41580 from: 1 to: 495

```

2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
103 TCTGGCTGCCCTTATTGGAGATGTGATTTCCAGACAAATCAATCCACA 152
18 pIleSerIleProGluTyrLysGluLeuGluGlnGluPheIleAspSerA 35
153 AGTGTCTTAAGACTGAATACAGAACTTCTTCAAGAGATTCATAGACGACA 202
35 spAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
203 ATGCCACTACAAATGCCATAGATGAATTTGAAGGAATGTTTCTTAACCAA 252
52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
253 ACGGATGAACACTCGACCAATGTTGAGGTGTTTATGCAATTAATATATGA 302
68 pSerIleTrpCysAsnLeu 74
303 CAGCAGTCTTTGTGATTTA 321

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT:T50925
seq_documentation_block:
ID T50925 standard; CDNA; 503 BP.
XX AC T50925;
XX DT 12-AUG-1997 (first entry)
XX DE cDNA encoding mammary-specific secretory protein, mammaglobin.
XX KW mammaglobin; mammary-secretory protein; breast cancer; detection;
XX KW neoplastic disease; diagnosis; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT 5'UTR 1..60
XX /tag= a
XX CDS 61..342
XX /tag= b
XX /product= mammaglobin
XX sig_peptide 61..117
XX /tag= c
XX mat_peptide 118..339
XX /tag= d
XX 3'UTR 343..503
XX /tag= e
XX polyA_signal 489..494
XX /tag= f
XX misc_feature 1..403
XX /tag= g
XX /note= "403 bp fragment isolated by RACE PCR"
XX /tag= h
XX /note= "DEST002"
XX WO9638463-A1.
XX PN 05-DEC-1996.
XX PF 31-MAY-1996; 96WO-US08235.
XX PR 31-MAY-1995; 95US-0455896.
XX PA (UNIW ) UNIV WASHINGTON.
XX PI Fleming TP, Watson MA;
XX WPI: 1997-034299/03.
XX P-PSDB; W10179.

```

```

XX PT Nucleic acid encoding mammary-specific secretory protein,
XX PT mammaglobin - used to develop prods. for the early diagnosis and
XX PT treatment of breast cancer neoplastic disease
XX PS Claim 1; Fig 2; 54pp; English.
XX CC The present sequence encodes a mammary-specific secretory protein
XX CC designated mammaglobin, which is overexpressed in 27% of stage I primary
XX CC breast cancer tumours. The anonymous sequence tag previously designated
XX CC DES1002 was used to demonstrate that the gene product mammaglobin is
XX CC abundant in the breast cancer tumour cell line MDA-MB-415. To isolate
XX CC the full-length mammaglobin cDNA, the mRNA was reverse transcribed from
XX CC this cell line and cloned using the RACE PCR technique. The nucleic acid
XX CC and protein can be used to develop prods. e.g. antibodies or probes, for
XX CC the detection and treatment of breast neoplastic disease.
XX CC Sequence 503 BP; 146 A; 118 C; 97 G; 142 T; 0 other;
XX CC
XX CC alignment_scores:
XX CC   Quality: 208.00   Length: 73
XX CC   Ratio: 3.355     Gaps: 0
XX CC Percent Similarity: 84.932   Percent Identity: 53.425
XX CC
XX CC alignment_block:
XX CC US-09-367-009-3 x T50925
XX CC
XX CC align seg 1/1 to: T50925 from: 1 to: 503
XX CC
XX CC 2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
XX CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX CC 118 TCTGGCTGCCCTTATTGGAGAATGTGATTTCCAAAGACAATCAATCCACA 167
XX CC
XX CC 18 pIleSerIleProGluTyrLysGluLeuGlnGluPheIleAspSera 35
XX CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX CC 168 AGTGCTCTAAGACTGAATACAAAGAACTTCTTCAAGAGTTCATAGACGACA 217
XX CC
XX CC 35 spAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
XX CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX CC 218 ATGCCACTACAAATGCCATAGATGAATGAAGGATGTTTCTTAACCAA 267
XX CC
XX CC 52 SerHisArgThrLeuLysAsnPheGlyLeuMetMethIsthValTyrAs 68
XX CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX CC 268 ACGGATGAACACTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATATGA 317
XX CC
XX CC 68 pSerIleTrpCysAsnLeu 74
XX CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX CC 318 CAGCAGTCTTTTGTGATTGA 336
XX CC
XX CC seq_name: /SIDS2/gcdata/geneseq/geneseq/NA1998.DAT:V17905
XX CC
XX CC seq_documentation_block:
XX CC ID V17905 standard; DNA; 503 BP.
XX CC
XX CC V17905;
XX CC
XX CC 13-JUL-1998 (first entry)
XX CC
XX CC Mammaglobin gene SEQ ID NO:1.
XX CC
XX CC Mammaglobin; detection; diagnosis; breast cancer; tumour; antibody;
XX CC gene therapy; human; ss.
XX CC
XX CC Homo sapiens.
XX CC
XX CC Key Location/Qualifiers
XX CC CDS 61..342
XX CC /*tag= a
XX CC /product= "mammaglobin"
XX CC
XX CC W09807753-A1.

```

```

PD 26-FEB-1998.
XX 19-AUG-1997; 97WO-US14666.
XX 15-AUG-1997; 97US-0697106.
XX 19-AUG-1996; 96US-0697106.
XX (ABBO ) ABBOTT LAB.
XX Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN;
XX Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD;
XX Roberts-Rapp L, Russell JC, Stroupe SD;
XX WPI; 1998-169096/15.
XX P-PSDB; W48432.
XX Antibodies to mammaglobin polypeptide(s) - used for detecting,
XX diagnosing, preventing or treating diseases or conditions of breast
XX such as breast cancer
XX
XX Example 4; Page 89; 105pp; English.
XX
XX The present sequence encodes mammaglobin which is used in an example of
XX the present invention. The present invention describes an antibody (A)
XX which specifically binds to at least 1 mammaglobin epitope (ME) which is
XX derived from an amino acid sequence having at least 50% identity to an
XX amino acid sequence (see W48432) and fragments. Also described are: (1)
XX an assay kit for determining the presence of mammaglobin antigen (MA) in
XX a test sample, comprising a container containing an antibody as in (A);
XX (2) a method for producing antibodies which specifically bind to a MA,
XX comprising administering an isolated immunogenic polypeptide or fragment
XX to elicit an immune response, where the immunogenic polypeptide
XX comprises at least 1 ME and has at least 50% identity to a sequence
XX (see W48432) and fragments, and (3) a method for producing antibodies
XX which specifically bind to a MA comprising administering to a mammal a
XX plasmid comprising a sequence which encodes at least 1 ME derived from a
XX polypeptide having an amino acid sequence (see W48432) and fragments.
XX The products and methods can be used for detecting, diagnosing, staging,
XX monitoring, prognosticating, preventing or treating, or determining
XX predisposition to diseases or conditions of the breast such as breast
XX cancer.
XX
XX SQ Sequence 503 BP; 146 A; 118 C; 97 G; 142 T; 0 other;

```

```

alignment_scores:
  Quality: 208.00   Length: 73
  Ratio: 3.355     Gaps: 0
Percent Similarity: 84.932   Percent Identity: 53.425

alignment_block:
US-09-367-009-3 x V17905
..
Align seg 1/1 to: V17905 from: 1 to: 503

2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
118 TCTGGCTGCCCTTATTGGAGAATGTGATTTCCAAAGACAATCAATCCACA 167

18 pIleSerIleProGluTyrLysGluLeuGlnGluPheIleAspSera 35
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
168 AGTGCTCTAAGACTGAATACAAAGAACTTCTTCAAGAGTTCATAGACGACA 217

35 spAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
218 ATGCCACTACAAATGCCATAGATGAATGAAGGATGTTTCTTAACCAA 267

52 SerHisArgThrLeuLysAsnPheGlyLeuMetMethIsthValTyrAs 68
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
268 ACGGATGAACACTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATATGA 317

68 pSerIleTrpCysAsnLeu 74
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```


Percent Similarity: 84.932 Percent Identity: 53.425

alignment_block:

US-09-367-009-3 x A12632

Align seq 1/1 to: A12632 from: 1 to: 503

```

2  SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
   |||||
1118 TCTGCTGCCCTTTATTGGAGAATGTGATTTCAAGACCAATCAATCCACA 167
   |||||

18 pIeserIleProGluTrvLysGluLeuLeuGlnGluPheIleAspSerA 35
   |||||
168 AGTGCTCAAGACTCAATACAAAGAACTCTTCAAGAGTTCATAGACGACA 217
   |||||

35 spAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
   |||||
218 ATGCCACTACAAATGCCATAGATGAATTAAGAGAAATGTTTCTTAACCAA 267
   |||||

52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTrAs 68
   |||||
268 ACGNGTGAACCTCTGACCAATGTGAGGTGTTTATGCAATTAATATCA 317
   |||||

68 pSerIleTrpCysAsnLeu 74
   |||||
318 CAGCAGCTCTTTGTGATTTA 336

```

seq_name: /SIDS2/qcadata/qeneseq/qeneseqn/NA2001.DAT:F17694

seq_documentation_block:

F17694 standard; cDNA; 511 BP.

F17694;

13-MAR-2001 (first entry)

Human breast cancer associated 19439-1 coding sequence.

Human: breast cancer associated gene; vaccine; diagnosis; therapy; ss.

Homo sapiens.

W0200060076-A2

12-OCT-2008

15-FEB-2000: 2000WO-11505308

02-APR-1999. 9915-0285480

02-APR-1999; 99US-0283480.
23-JUN-1999; 99US-0339338.

02-SEP-1999; 99US-0389681.

03-NOV-1999; 99US-0433826.

(CORI-) CORIXA CORP.

Yuqiu J, Dillon DC, Mitcham JL, Xu J, Harlocker SL;

WPI; 2001-122627/13.

An isolated polypeptide useful for the treatment and diagnosis of tumors e.g. breast cancer comprises at least an immunogenic portion of a breast tumor protein -

Claim 66: page 146: 238pp: English.

The present invention provides the coding sequences and some protein sequences of proteins associated with breast cancer in humans. These sequences can be used in the diagnosis and treatment of cancers, particularly breast tumours.

Sequence 511 BP: 143 A: 98 C: 115 G: 155 T: 0 other:

[illegible]

OM of: US-09-367-009-3 to: EST:* out_format : pfs
 Date: Apr 28, 2001 4:21 PM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
 -MODEL=framet p2n.model -DEV=xlp
 -O=/cpn2.1/USPTC_spool/US09367009/runat_24042001_151806_19009/app_query.fasta_1.131
 -DB=EST -QFMT=fastap -SUFFIX=first -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=biosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR SCORE=oct -THR MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=200000000 -USER=US09367009 -CGN1_1_2610 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
 Query: US-09-367-009-3
 Query length: 74
 Database: EST:*
 Database sequences: 9623517
 Database length: 73081774
 Search time (sec): 1128.660000

score_list:
 Sequence Strd Orig zScore EScore Len Documentation
 gb_est69:BE044893 + 390.00 949.88 8.1e-44 351 BE044893 hn08c04.x1 NCI_CGAP_Th
 gb_est69:BE044895 + 390.00 949.48 8.6e-44 365 BE044895 hn08d03.x1 NCI_CGAP_Th
 gb_est5:AA297402 + 390.00 948.66 9.5e-44 396 AA297402 EST112936 Endometrial
 gb_est7:AA398560 + 390.00 947.71 1.1e-43 435 AA398560 zt73f02.s1 Soares_test
 gb_est8:AA493295 + 390.00 947.11 1.1e-43 435 AA493295 ng97e06.s1 NCI_CGAP_Th
 gb_est21:AI491987 + 390.00 947.18 1.1e-43 458 AI491987 to07f11.x1 NCI_CGAP_Th
 gb_est23:AI659370 + 390.00 947.14 1.2e-43 460 AI659370 tullc09.x1 NCI_CGAP_Th
 gb_est8:AA393164 + 390.00 946.48 1.3e-43 491 AA393164 zt73f02.rl Soares_test
 gb_est46:AW451131 + 390.00 946.42 1.3e-43 494 AW451131 UI-H-B13-alg-h-05-0-UI
 gb_est53:AW966513 + 390.00 943.87 1.8e-43 636 AW966513 EST378587 MAGE resequ
 gb_est53:AW966509 + 390.00 943.69 1.8e-43 647 AW966509 EST378583 MAGE resequ
 gb_est51:AW779377 + 385.00 934.87 5.6e-43 458 AW779377 nh79g08.x1 NCI_CGAP_K1
 gb_est27:AI936084 + 385.00 934.59 5.8e-43 471 AI936084 wo61d06.x1 NCI_CGAP_Th
 gb_est5:AA298750 + 384.00 936.10 4.8e-43 318 AA298750 EST114585 Uterus Homo
 gb_est25:AI800231 + 384.00 932.28 7.8e-43 464 AI800231 ti75c07.x1 NCI_CGAP_K1
 gb_est43:AW523774 + 379.00 921.02 3.3e-42 418 AW523774 xo56b10.x1 NCI_CGAP_Th
 gb_est47:AW510561 + 372.00 880.76 5.8e-40 357 AW510561 xo37g03.x1 NCI_CGAP_Th
 gb_est53:AW966498 + 362.00 853.66 1.9e-38 456 AW966498 EST378572 MAGE resequ
 gb_est8:AA525178 + 289.00 700.70 6.2e-30 368 AA525178 n151C09.s1 NCI_CGAP_Th
 gb_est91:BF745480 + 288.00 704.55 3.8e-30 197 BF745480 CM2-BT0826-021100-459-
 gb_est5:AA297452 + 271.00 659.58 1.2e-27 268 AA297452 EST112981 Endometrial
 gb_est5:AA297456 + 267.00 650.31 3.9e-27 253 AA297456 EST112985 Endometrial
 gb_est27:AI937042 + 256.00 626.45 8.4e-26 184 AI937042 wp72d06.x1 NCI_CGAP_Th
 gb_est5:AA297405 + 244.00 594.13 5.3e-24 242 AA297405 EST112939 Endometrial
 gb_est16:AI124690 + 218.00 524.51 4.0e-20 421 AI124690 am60f04.x1 Johnston fr
 gb_est26:AI904282 + 208.00 501.14 8.0e-19 372 AI904282 PM-BT046-160199-249 BT
 gb_est42:AW175618 + 208.00 500.73 8.5e-19 388 AW175618 RC3-BT0046-310899-002-
 gb_est24:AI732534 + 208.00 499.53 9.9e-19 436 AI732534 nh79e11.x5 NCI_CGAP_Th
 gb_est14:AA991451 + 208.00 499.35 1.0e-18 444 AA991451 os52f11.s1 NCI_CGAP_Th
 gb_est8:AA513640 + 208.00 499.28 1.0e-18 447 AA513640 nh79e11.s1 NCI_CGAP_Th
 gb_est25:AI791322 + 208.00 498.93 1.1e-18 463 AI791322 aes4e06.y5 Stratagene
 gb_est26:AI904278 + 203.00 488.74 3.9e-18 375 AI904278 PM-BT046-160199-254 BT
 gb_est53:AW972474 + 198.00 473.32 2.8e-17 510 AW972474 EST384565 MAGE resequ
 gb_est8:AA513647 + 186.00 446.44 9.0e-16 390 AA513647 nh79f09.s1 NCI_CGAP_Th
 gb_est15:AI014324 + 180.00 438.98 2.3e-15 192 AI014324 am51b10.s1 Johnston fr
 gb_est26:AI905265 + 176.00 425.85 1.3e-14 263 AI905265 RC-BT081-260499-022 BT
 gb_est5:AA297482 + 175.00 426.73 1.1e-14 189 AA297482 EST113009 Endometrial
 gb_est42:AW175603 + 170.00 410.37 9.1e-14 282 AW175603 RC3-BT0046-310899-002-
 gb_est26:AI905273 + 169.00 407.48 1.3e-13 294 AI905273 RC-BT081-260499-003 BT
 gb_est72:BE329311 + 163.00 388.18 1.6e-12 460 BE329311 hq39c06.x1 NCI_CGAP_Th
 gb_est72:BE329440 + 163.00 388.18 1.6e-12 460 BE329440 hq37d04.x1 NCI_CGAP_Th

gb_est87:BF417380 - 163.00 387.99 1.6e-12 469 BF417380 UI-R-CNO-bld-e-05-0
 gb_est87:BF417713 - 163.00 387.97 1.6e-12 470 BF417713 UI-R-CNO-bld-c-07-0
 gb_est87:BF417790 - 163.00 387.92 1.6e-12 472 BF417790 UI-R-CNO-bld-c-01-0
 gb_est86:BF411207 - 163.00 387.75 1.7e-12 480 BF411207 UI-R-CNO-bld-d-03-0

seq_name: gb_est69:BE044893

seq_documentation_block: 351 bp mRNA EST 08-JUN-2000
 LOCUS BE044893
 DEFINITION hn08c04.x1 NCI_CGAP_Thy6 Homo sapiens cDNA clone IMAGE:3021510.3,
 similar to TR:075556 075556 MAMMAGLOBIN B PRECURSOR. ;, mRNA
 sequence.
 ACCESSION BE044893
 VERSION BE044893
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 351)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Klaus Kaserer M.D., Bruno Niederle M.D., Mike
 Emmert-Buck M.D. Ph.D., Vlado Knezevic M.D.
 cDNA Library Preparation: Krizman Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL, send email to:
 info@image.llnl.gov
 Possible reversed clone: similarity on wrong strand
 Possible reversed clone: polyT not found
 Seq primer: -40UP from Gibco
 High quality sequence stop: 351.

FEATURES
 Location/Qualifiers
 1..351
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="3021510"
 /clone_lib="NCI_CGAP_Thy6"
 /tissue_type="normal epithelium"
 /lab_host="DH10B"
 /note="Organ: thyroid; Vector: pAMP10; mRNA made from
 normal thyroid epithelium, cDNA made by oligo-dT priming.
 Non-directionally cloned into UDG sites. Size-selected on
 agarose gel, average insert size 500 bp. Primary library.
 cDNA Library Preparation: David B. Krizman, Ph.D.
 REFERENCE: Krizman et al. (1996) Cancer Research
 56:5380-5383."

BASE COUNT 98 a 81 c 83 g 89 t
 ORIGIN
 alignment_scores:
 Quality: 390.00 Length: 74
 Ratio: 5.270 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 98.649

alignment_block:
 US-09-367-009-3 x BE044893
 Align seg 1/1 to: BE044893 from: 1 to: 351
 1 AspSerGlyCysLysLeuLeuGluAspMetValcLulysThrIleAsnSe 17
 ||||||||||||||||||||||||||||||||||||||||||||||||
 59 GATCTCGCTGCAACTCCCTGGAGCATGTTGAAAGACCAATTC 108
 17 rAspIleSerIleProGluTyrLysGluLeuGlnGluPheIleAsps 34

```

|||||
109 CGACATATCTATACCTGAATCAAGAGCTTCTTCAAGAGTTTCATAGACA 158
|||||
34 erAspAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
|||||
159 GTGATGCCGCTGCAGAGGCTATGGGAAATTCAGAGCAGTGTTCCTCAAC 208
|||||
51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValty 67
|||||
209 CAGTCACATAGAACTCGAAAACCTTGGACTGATGATGATGATGATGATGTA 258
|||||
67 rAspSerIleTrpCysAsnLeu 74
|||||
259 CGACAGCATTTGGTGTATATG 280
|||||
seq_name: gb_est69:BE044895

seq_documentation_block: 365 bp mRNA 08-JUN-2000
LOCUS BE044895
DEFINITION hn08d03.x1 NCI-CGAP_Thy6 Homo sapiens cDNA clone IMAGE:3021509 3'
similar to TR:075556 O75556 MAMMAGLOBIN B PRECURSOR. ; mRNA
sequence.
ACCESSION BE044895
VERSION BE044895.1 GI:8361948
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 365)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Klaus Kaserer M.D., Bruno Niederle M.D., Mike
Emmert-Buck M.D. Ph.D., Vlado Knezevic M.D.
CDNA Library Preparation: Krizman Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Possible reversed clone: similarity on wrong strand
Possible reversed clone: polyt not found
Seq primer: -400P from Glibco
High quality sequence stop: 364.
FEATURES
Location/Qualifiers
1..365
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3021509"
/clone_lib="NCI-CGAP_Thy6"
/tissue_type="normal epithelium"
/lab_host="DH108"
/note="Organ: thyroid; Vector: pAMP10; mRNA made from
normal thyroid epithelium, cDNA made by oligo-dt priming.
Non-directionally cloned into UDG sites. Size-selected on
agarose gel, average insert size 500 bp. Primary library.
cDNA Library Preparation: David B. Krizman, Ph.D.
REFERENCE: Krizman et al. (1996) Cancer Research
56:5380-5383."
103 a 85 c 85 g 92 t

```

alignment_scores:

Quality:	390.00	Length:	74
Ratio:	5.270	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	98.649

```

alignment_block:
US-09-367-009-3 x BE044895
Align seg 1/1 to: BE044895 from: 1 to: 365
1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
|||||
59 GATTCGGCTGCAAACTCCCTGGAGGACATGCTTGAAGAAGACCATCAATC 108
|||||
17 rAspIleSerIleProGluTrpLysGluLeuLeuGlnGluPheIleAsps 34
|||||
109 CGACATATCTATACCTGAATCAAGAGCTTCTTCAAGAGTTTCATAGACA 158
|||||
34 erAspAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
|||||
159 GTGATGCCGCTGCAGAGGCTATGGGAAATTCAGAGCAGTGTTCCTCAAC 208
|||||
51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValty 67
|||||
209 CAGTCACATAGAACTCGAAAACCTTGGACTGATGATGATGATGATGATGTA 258
|||||
67 rAspSerIleTrpCysAsnLeu 74
|||||
259 CGACAGCATTTGGTGTATATG 280
|||||
seq_name: gb_est5:AA297402

seq_documentation_block: 396 bp mRNA 18-APR-1997
LOCUS AA297402
DEFINITION EST112936 Endometrial tumor Homo sapiens cDNA 5' end similar to
similar to steroid-binding protein, C3 chain, prostate, mRNA
sequence.
ACCESSION AA297402
VERSION AA297402.1 GI:1949735
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 396)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
,C.J., Lee,N.H., Kirkness,E.F., Weinstein,K.G., Gocayne,J.D., White
,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodak,A.,
Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M.,
Kelley,J.C., Liu,L.-I., Marmaros,S.N., Merrick,J.M., Peltigri,S.M.,
Moreno-Palanco,R.F., McDonald,L.A., Nguyen,D.T., Saudek,D.M., Shirley,R.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.F., Ferris,A., Fischer,C., Hastings,G.A., He,W.W.,
Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H., Raymond,L.,
Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon
,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
Venter,J.C.
TITLE Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE 96026280
COMMENT Other ESTs: THC168663
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi.html)
Seq primer: M13 Reverse.

```

```

FEATURES
  source
    Location/Qualifiers
      1..396
        /organism="Homo sapiens"
        /db_xref="ATCC (inhost):138082"
        /db_xref="taxon:9606"
        /clone_lib="Endometrial tumor"
        /sex="female"
        /dev_stage="adult"
        /note="Organ: endometrium; Vector: pBluescript SK-;
        Site_1: EcoRI; Site_2: XhoI"
      93 g
      96 t
      1 others

BASE COUNT
  111 a
  95 c
  93 g
  96 t
  1 others

alignment_scores:
  Quality: 390.00 Length: 74
  Ratio: 5.270 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 98.649

alignment_block:
  US-09-367-009-3 x AA297402
  ..

  Align seg 1/1 to: AA297402 from: 1 to: 396
    1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
    92 GATTCTGGCTGCAAACTCTCGGAGGACATGGTTGAAAAGACCATCAATTC 141
    17 rAspIleSerIleProGluTyrLysGluLeuGluGlnGluPheIleAspS 34
    142 CGACATATCTATACCTGTAATACAAAGAGCTTCTTCAAGAGTTCATAGACA 191
    34 erAspAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
    192 GTGATCCGCTCGAGAGGTATGGGAAATTCGAAGCAGTGTTCCTCAAC 241
    51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValty 67
    242 CAGTCACATAGAACTCTGAAAACCTTTGGAGTATGATGATGATGATGATGTA 291
    67 rAspSerIleTrpCysAsnLeu 74
    292 CGACAGCAATTTGGTGAATATG 313

  seq_name: gb_est7:AA398560

seq_documentation_block:
  LOCUS AA398560 435 bp mRNA EST 16-MAY-1997
  DEFINITION 273f02.s1 Soares testis_NHT Homo sapiens cDNA clone IMAGE:727995
  3' similar to TR:G1199596 G1199596 MAMMAGLOBIN. ; mRNA sequence.
  ACCESSION AA398560
  VERSION AA398560.1 GI:2051669
  KEYWORDS EST.
  SOURCE human.
  ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE 1 (bases 1 to 435)
  AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
    Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
    Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie,
    T., Waterston, R., and Wilson, R.
  TITLE WashU-Wetck Est Project 1997
  JOURNAL Unpublished (1997)
  COMMENT Contact: Wilson RK
    Washington University School of Medicine
    444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
    Tel: 314 286 1800
    Fax: 314 286 1810
    Email: est@watson.wustl.edu
    This clone is available royalty-free through LNL ; contact the
    IMAGE Consortium (info@image.lni.gov) for further information.
    Seq primer: -41m13 fwd. Ef from Amersham.

FEATURES
  source
    Location/Qualifiers
      1..435
        /organism="Homo sapiens"
        /db_xref="GDB:5924907"
        /db_xref="taxon:9606"
        /clone_lib="IMAGE:727995"
        /clone_lib="Soares_testis_NHT"
        /sex="male"
        /lab_host="DH10B"
        /note="Vector: p773D-Pac (Pharmacia) with a modified
        polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
        was prepared from mRNA obtained from Clontech Laboratories
        , Inc. and primed with a Not I - oligo(dT) primer {5'
        TCTTACCACATCTGAAGTGGAGCGCGCCCAATTTTCTTTT 3'}.
        Double-stranded cDNA was ligated to Eco RI adaptors
        (Pharmacia), digested with Not I and cloned into the Not I
        and Eco RI sites of the modified p773 vector. Library
        went through one round of normalization to Cot5, and was
        constructed by Bento Soares and M. Fatima Bonaldo. "
      124 a
      87 c
      95 g
      129 t

BASE COUNT
  124 a
  87 c
  95 g
  129 t

alignment_scores:
  Quality: 390.00 Length: 74
  Ratio: 5.270 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 98.649

alignment_block:
  US-09-367-009-3 x AA398560/rev
  ..

  Align seg 1/1 to reverse of: AA398560 from: 1 to: 435
    1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
    406 GATTCTGGCTGCAAACTCTCGGAGGACATGGTTGAAAAGACCATCAATTC 357
    17 rAspIleSerIleProGluTyrLysGluLeuGluGlnGluPheIleAspS 34
    356 CGACATATCTATACCTGTAATACAAAGAGCTTCTTCAAGAGTTCATAGACA 307
    34 erAspAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
    306 GTGATCCGCTCGAGAGGTATGGGAAATTCGAAGCAGTGTTCCTCAAC 257
    51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValty 67
    256 CAGTCACATAGAACTCTGAAAACCTTTGGAGTATGATGATGATGATGATGTA 207
    67 rAspSerIleTrpCysAsnLeu 74
    206 CGACAGCAATTTGGTGAATATG 185

  seq_name: gb_est8:AA493295

seq_documentation_block:
  LOCUS AA493295 435 bp mRNA EST 19-AUG-1997
  DEFINITION ng97e06.s1 NCI-CCAP_Thyl Homo sapiens cDNA clone IMAGE:942754
  similar to TR:G1199596 G1199596 MAMMAGLOBIN. ; mRNA sequence.
  ACCESSION AA493295
  VERSION AA493295.1 GI:2223136
  KEYWORDS EST.
  SOURCE human.
  ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE 1 (bases 1 to 435)
  AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
    Tumor Gene Index
  JOURNAL Unpublished (1997)
  COMMENT Contact: Robert Strausberg, Ph.D.
    Tel: (301) 496-1550

```

Email: Robert.Strausberg@nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 887 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 401.

Location/Qualifiers
1. 435
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:942754"
/clone_lib="NCI-CGAP_Thyl"
/tissue_type="thyroid"
/lab_host="DH10B"
/note="Vector: pAMP10; mRNA made from invasive thyroid tumor, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Reference: Krizman et al. (1996) Cancer Research 56:5380-5383."
BASE COUNT 122 a 86 c 94 g 133 t
ORIGIN

alignment_scores:
Quality: 390.00 Length: 74
Ratio: 5.270 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.649
alignment_block:
US-09-367-009-3 x AA493295/rev ..

Align seg 1/1 to reverse of: AA493295 from: 1 to: 435
1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
408 GATTCTGGCTGCAAACTCCCTGGAGGACATGTTGAAAAGACCATCAATTC 359
17 rAspSerIleProGluTyLysGluLeuLeuGlnGluPheIleAsps 34
358 CGACATATCTATACCTGAAATACAAGAGCTTCTTCAAGAGTTCATAGACA 309
34 erAspAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
308 GTGATGCGCTGCAGAGCTATGGGAAATTCAGAGCAGTGTTCCTCAAC 259
51 GlnSerHisArgThrLeuLysAspPheGlyLeuMetMetHisThrValTy 67
258 CAGTCACATAGAACTCTGAAAACCTTTGGAGTGTGATGCATACAGTGTA 209
67 rAspSerIleTrpCysAsnLeu 74
208 CGACAGCATTTGGTGTATATG 187

seq_name: gb_est21:AI491987
seq_documentation_block:
LOCUS AI491987 458 bp mRNA EST 12-MAY-1999
DEFINITION to07f11.x1 NCI-CGAP_022 Homo sapiens cDNA clone IMAGE:2178381 3' similar to SW:MAMG_HUMAN Q13296 MAMMAGLOBIN PRECURSOR. [1]; mRNA sequence.
ACCESSION AI491987.1 GI:4393001
VERSION AI491987
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 458)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 577 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 399.
POLYA=No.

FEATURES
source

Location/Qualifiers
1. 458
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2178381"
/clone_lib="NCI-CGAP_022"
/tissue_type="moderately-differentiated endometrial adenocarcinoma, 3 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.85 kb. Life Technologies catalog #:
11539-012"

BASE COUNT 132 a 93 c 101 g 132 t
ORIGIN

alignment_scores:
Quality: 390.00 Length: 74
Ratio: 5.270 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.649
alignment_block:
US-09-367-009-3 x AI491987/rev ..

Align seg 1/1 to reverse of: AI491987 from: 1 to: 458
1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
404 GATTCTGGCTGCAAACTCCCTGGAGGACATGTTGAAAAGACCATCAATTC 355
17 rAspSerIleProGluTyLysGluLeuLeuGlnGluPheIleAsps 34
354 CGACATATCTATACCTGAAATACAAGAGCTTCTTCAAGAGTTCATAGACA 305
34 erAspAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
304 GTGATGCGCTGCAGAGCTATGGGAAATTCAGAGCAGTGTTCCTCAAC 255
51 GlnSerHisArgThrLeuLysAspPheGlyLeuMetMetHisThrValTy 67
254 CAGTCACATAGAACTCTGAAAACCTTTGGAGTGTGATGCATACAGTGTA 205
67 rAspSerIleTrpCysAsnLeu 74
204 CGACAGCATTTGGTGTATATG 183

seq_name: gb_est23:AI659370
seq_documentation_block:
LOCUS AI659370 460 bp mRNA EST 15-DEC-1999
DEFINITION tucic09.x1 NCI-CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2250736 3' similar to TR:O75556 O75556 MAMMAGLOBIN B PRECURSOR. ;, mRNA


```

sequence.
ACCESSION   AI659370
VERSION     AI659370.1  GI:4762940
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 460)
            NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            DNA Sequencing by: Greg Lennon, Ph.D.
            Clone distribution: NCI-CCAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 559 Std Error: 0.00
            Seq primer: -400P from Gibco
            High quality sequence stop: 448.
FEATURES             Location/Qualifiers
     source           1..460
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NCI-CCAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI-CCAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     985608-986759, 1101192-1101959, and 1217928-1220615)."
     BASE COUNT      132 a  94 c  101 g  133 t
     ORIGIN
132 a  94 c  101 g  133 t
                                     ..
Align seg 1/1 to reverse of: AI659370 from: 1 to: 460
1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17'
|||||
406 GATTCGGCTGCAAACTCCGGAGGACATGGTTGAAAAGACCATCAATTC 357
|||||
17 rAspIleSerIleProGluTyrLysGluLeuLeuGlnGluPheIleAsps 34
|||||
356 CGACATATCTATACCTGAATACAAAGAGCTCTCTCAAGAGTTCATGACA 307
|||||
34 erAspIlaAlaAlaGluAlaMeGlyLysPheLysGlnCysPheLeuAsn 50
|||||
306 GTGATCCGCTGCAGAGGCTATGGGAAATTCAGGAGTGTTCCTCAAC 257
|||||
51 GlnSerHisArgThrIleLysAsnPheGlyLeuMetMethIshThrValty 67
|||||

```

```

256 CAGTCACATAGAACTCTGAAAAAATTTGGACTGATGATGCATACAGTGTA 207
67 rAspSerIleIleTrpCysAsnLeu 74
|||||
206 CGACAGCATTTGGTGAATATG 185
seq_name: gb_est6:AA393164
seq_documentation_block:
LOCUS       AA393164               491 bp    mRNA                    EST
DEFINITION  zt73f02.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:727995
5' similar to TR:G1199596 G1199596 MAMMAGLOBIN.; mRNA sequence.
ACCESSION   AA393164
VERSION     AA393164.1  GI:2046134
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 491)
            Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
            Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
            Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie
            , T., Waterston, R. and Wilson, R.
            WashU-Merck EST Project 1997
            Unpublished (1997)
            Contact: Wilson RK
            Washington University School of Medicine
            444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28ml3 rev2 ET from Amersham
            High quality sequence stop: 380.
FEATURES             Location/Qualifiers
     source           1..491
                     /organism="Homo sapiens"
                     /db_xref="GDB:5924907"
                     /db_xref="taxon:9606"
                     /clone_lib="IMAGE:727995"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT73D-Pac (Pharmacia) with a modified
                     polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech Laboratories
                     Inc., and primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGGCGGAGCGGCCGCCCAATTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT73 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
     BASE COUNT      149 a  109 c  100 g  133 t
     ORIGIN
149 a  109 c  100 g  133 t
                                     ..
Align seg 1/1 to: AA393164 from: 1 to: 491
1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
|||||
73 GATTCGGCTGCAAACTCCGGAGGACATGGTTGAAAAGACCATCAATTC 122

```

pool of 3,840 arrayed clones from NCI_CGAP_Sub1 (IMAGE Clones 2708616-2710535) and NCI_CGAP_Sub2 (IMAGE Clones 2710536-2712455) (10% of the driver population), plus a pool of 11,136 clones from NCI_CGAP_Sub3 (IMAGE Clones 2712456-2723591) (10% of the driver population), plus a pool of 5,472 clones from NCI_CGAP_Sub4 (IMAGE Clones 2723592-2728969) (70% of the driver population). Subtraction was performed as previously described [Bonaldi, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6, 791-806].

TAG_LIB=NCI_CGAP_Br2
TAG_TISSUE=breast
TAG_SEQ=AAACC

BASE COUNT 132 a 109 g 154 t 1 others
ORIGIN

alignment_scores:
Quality: 390.00 Length: 74
Ratio: 5.270 Gaps: 0
Percent Similarity: 100.00 Percent Identity: 98.649

alignment_block:
US-09-367-009-3 x AW451131/rev ..

Align seg 1/1 to reverse of: AW451131 from: 1 to: 494

1 AspSerGlyCysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
|||||
417 GATTCTCGCTGCAACCTCTGGAGGACATGTTGAAAGACCATCAATTC 368
|||||

17 rAspIleSerIleProGluTyrLysGluLeuLeuGlnGluPheIleAsps 34
|||||

367 CGACATATCTATACCTGATACAAAGAGCTCTTCAAGAGTTCATAGACA 318
|||||

34 erAspAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
|||||

317 GTGATGCCGCTGCAGAGCGCTATGGGAAATTCAGACGAGTGTTCCTCAAC 268
|||||

51 GlnSerHisArgThrLeuLysAsnPhcGlyLeuMetHisThrValty 67
|||||

267 CAGTCACATAGACTCTGAAAACTTTGGACTGATGATGCATACAGTGTA 218
|||||

67 rAspSerIleTyrCysAsnLeu 74
|||||

217 CGACAGCATTTGGTGTATATG 196
|||||

seq_name: gb_est53:AW966513

seq_documentation_block:
LOCUS AW966513 636 bp mRNA EST 01-JUN-2000
DEFINITION EST378587 MAGE resequences, MAGI Homo sapiens CDNA, mRNA sequence.
ACCESSION AW966513
VERSION AW966513.1 GI:8156349
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 636)
Hegd,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and Quackenbush,J.
Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray
Unpublished (2000)
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnqu@tigr.org

17 rAspIleSerIleProGluTyrLysGluLeuLeuGlnGluPheIleAsps 34
|||||

123 CGACATATCTATACCTGATACAAAGAGCTCTTCAAGAGTTCATAGACA 172
|||||

34 erAspAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
|||||

173 GTGATGCCGCTGCAGAGCGCTATGGGAAATTCAGACGAGTGTTCCTCAAC 222
|||||

51 GlnSerHisArgThrLeuLysAsnPhcGlyLeuMetHisThrValty 67
|||||

223 CAGTCACATAGACTCTGAAAACTTTGGACTGATGATGCATACAGTGTA 272
|||||

67 rAspSerIleTyrCysAsnLeu 74
|||||

273 CGACAGCATTTGGTGTATATG 294
|||||

seq_name: gb_est46:AW451131

seq_documentation_block:
LOCUS AW451131 494 bp mRNA EST 17-FEB-2000
DEFINITION UI-H-B13-alg-h-05-0-UI-si NCI_CGAP_Sub5 Homo sapiens CDNA clone
IMAGE:2736992 3', mRNA sequence.
ACCESSION AW451131
VERSION AW451131.1 GI:6991907
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 494)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I. M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward
POLYA=yes.

Location/Qualifiers
1. .494
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2736992"
/lab_host="NCI_CGAP_Sub5"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; NCI_CGAP_Sub5
is a subtracted library derived from NCI_CGAP_Sub4. The
NCI_CGAP_Sub5 library had 3 million recombinants. A
single-stranded DNA preparation of NCI_CGAP_Sub4 was used
as a tracer in a subtractive hybridization with a driver
comprising: the IMAGE pool (NCI_CGAP_Kid3 pool 1 LLAM
3334-3337, 3682-3683, 3798-3803 (IMAGE Clones)
1323376-1323911, 1456008-1456775, 1500552-1502855);
NCI_CGAP_Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778
(IMAGE Clones 1323912-1325931, 1471368-1472903
1492104-1493255); NCI_CGAP_Lus pool 1 LLAM 3575-3582,
3851-3854 (IMAGE Clones 1414920-1417991, 1520904-1522439
); NCI_CGAP_GC4 pool 1 LLAM 3164-3167, 3716-3720,
3733-3735 (IMAGE Clones 1257096-1258631, 1469064-1470983,
1475592-1476743); NCI_CGAP_P222 pool 1 LLAM 2457-2459,
2758-2759, 3062-3066 (IMAGE Clones 985608-986759
1101192-1101959, 1217928-1220615); NCI_CGAP_Colo pool 1
LLAM 2644-2653, 2871-2872 (IMAGE Clones 1057416-1061255
1144584-1145351). (10% of the driver population), plus a

Plate: 229
 Seq primer: Forward.
 Location/Qualifiers
 1..636
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="MAGE resequences, MAGI"
 /note="vector: pBluescriptSKm"
 BASE COUNT 189 a 152 c 134 g 161 t
 ORIGIN

alignment_scores:
 Quality: 390.00 Length: 74
 Ratio: 5.270 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 98.649

alignment_block:
 US-09-367-009-3 x AW966513 ..
 Align seg 1/1 to: AW966513 from: 1 to: 636
 1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleasnse 17
 91 GATTCGGCTGCAAACTCCTGGAGGACATGGTTGAAAAGACCAATCAATC 140
 17 rAspIleSerIleProGluTyrLysGluLeuGluGlnGluPheIleAsps 34
 141 CGACATATCTATACCTGAATACAAGAGCTTCTTCAAGAGTTTCATAGACA 190
 34 erAspAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
 191 GTGATCGCGCTGCAGAGGCTATGGGAAATTCAGAGAGTGTTCCTCAAC 240
 51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMethIsthrValty 67
 241 CAGTCACATAGAACTCTGAAAACCTTTGGACTGATGATGATACAGTGTA 290
 67 rAspSerIleTyrCysAsnLeu 74
 291 CGACAGCATTTGGTGAATATG 312

seq_name: gb_est53:AW966509

seq_documentation_block:
 LOCUS AW966509 647 bp mRNA EST 01-JUN-2000
 DEFINITION EST378583 MAGE resequences, MAGI Homo sapiens cDNA, mRNA sequence.
 ACCESSION AW966509
 VERSION AW966509.1 GI:8156345
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 647)
 AUTHORS Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
 I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
 Quackenbush,J.
 TITLE Assessment of gene expression patterns in a model of colon tumor
 metastasis using a 19,200 element cDNA microarray
 JOURNAL Unpublished (2000)
 COMMENT Contact: John Quackenbush
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 3528
 Fax: 301 838 0208
 Email: johnq@tigr.org
 Plate: 229
 Seq primer: Forward.
 Location/Qualifiers
 1..647
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

/clone_lib="MAGE resequences, MAGI"
 /note="vector: pBluescriptSKm"
 BASE COUNT 193 a 155 c 140 g 159 t
 ORIGIN

alignment_scores:
 Quality: 390.00 Length: 74
 Ratio: 5.270 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 98.649

alignment_block:
 US-09-367-009-3 x AW966509 ..
 Align seg 1/1 to: AW966509 from: 1 to: 647
 1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleasnse 17
 91 GATTCGGCTGCAAACTCCTGGAGGACATGGTTGAAAAGACCAATCAATC 140
 17 rAspIleSerIleProGluTyrLysGluLeuGluGlnGluPheIleAsps 34
 141 CGACATATCTATACCTGAATACAAGAGCTTCTTCAAGAGTTTCATAGACA 190
 34 erAspAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
 191 GTGATCGCGCTGCAGAGGCTATGGGAAATTCAGAGAGTGTTCCTCAAC 240
 51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMethIsthrValty 67
 241 CAGTCACATAGAACTCTGAAAACCTTTGGACTGATGATGATACAGTGTA 290
 67 rAspSerIleTyrCysAsnLeu 74
 291 CGACAGCATTTGGTGAATATG 312

seq_name: gb_est51:AW779377

seq_documentation_block:
 LOCUS AW779377 458 bp mRNA EST 12-MAY-2000
 DEFINITION hn79g08.x1 NCI-CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3034142 3',
 similar to TR:O75556 O75556 MAMMAGLOBIN B PRECURSOR. ;, mRNA
 sequence.
 ACCESSION AW779377
 VERSION AW779377.1 GI:7793980
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 458)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL, send email to:
 info@image.llnl.gov
 Seq primer: -400P from Gibco
 High quality sequence stop: 458.
 Location/Qualifiers
 1..458
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3034142"

FEATURES
 source
 1..458
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3034142"

```

/clone_lib="NCI_CGAP_Kid11"
/lab_host="DH10B"
/notes="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(clone IDs 132376-132391, 145607-145675, and
1500552-1502853). Subtraction by Bento Soares and M.
Fatima Bonaldo."
BASE COUNT      132 a      99 g      132 t      1 others
ORIGIN
1  AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
|||||
404 GATTCTGGCTGCAAACTCTGGAGGACATGTTGAAAGACCATCAATTC 355
|||||
17 rAspIleSerIleProGluTyrLysGluLeuGluInclupheIleAsps 34
|||||
354 CGACATATCTATACCTGAATACAAAGAGCTTCTTCAAGAGTTGATAGACA 305
|||||
34 erAspAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
|||||
304 GTGATGCCGTGCAGAGGCTATGGGAAATTCGAAGCAGTGTTCCTCAAC 255
|||||
51 GlnSerHisArgThrLeuLysAsnPheGlyLysMetMetHisThrValty 67
|||||
254 CAGTCACATAGAACTCTGAAAACCTTGGAGTGTATGATGATGATAGTGA 205
|||||
67 rAspSerIleTrpCysAsnLeu 74
|||||
204 CGACAGCATTTGGTGTATATG 183
|||||
seq_name: gb_est27:AI936084

seq_documentation_block:      471 bp      mRNA      02-SEP-1999
LOCUS      AI936084
DEFINITION      w61d06.x1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:2459819 3'
similar to TR:075556 075556 MAMMAGLOBIN B PRECURSOR. ; mRNA
sequence.
ACCESSION      AI936084
VERSION      AI936084.1 GI:5674954
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 471)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center

/clone_lib="NCI_CGAP_Pr22"
/lab_host="DH10B"
/notes="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; 1st strand cDNA was prepared
from normal prostate bulk tissue, and was then primed with
a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT73 vector. Library is normalized, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      131 a      94 c      103 g      143 t
ORIGIN
1  AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
|||||
417 GATTCTGGCTGCAAACTCTGGAGGACATGTTGAAAGACCATCAATTC 368
|||||
17 rAspIleSerIleProGluTyrLysGluLeuGluInclupheIleAsps 34
|||||
367 CGACATATCTATACCTGAATACAAAGAGCTTCTTCAAGAGTTGATAGACA 318
|||||
34 erAspAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
|||||
317 GTGATGCCGTGCAGAGGCTATGGGAAATTCGAAGCAGTGTTCCTCAGC 268
|||||
51 GlnSerHisArgThrLeuLysAsnPheGlyLysMetMetHisThrValty 67
|||||
267 CAGTCACATAGAACTCTGAAAACCTTGGAGTGTATGATGATGATAGTGA 218
|||||
67 rAspSerIleTrpCysAsnLeu 74
|||||
217 CGACAGCATTTGGTGTATATG 196
|||||
seq_name: gb_est5:AA298750

seq_documentation_block:      318 bp      mRNA      18-APR-1997
LOCUS      AA298750
DEFINITION      EST114595 Uterus Homo sapiens cDNA 5' end similar to
steroid-binding protein, C3 chain, prostate, mRNA sequence.
ACCESSION      AA298750
VERSION      AA298750.1 GI:1951318
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 318)
AUTHORS      Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
,C.J., Lee,N.H., Kirkness,E.F., Weinscock,K.G., Gocayne,J.D., White
,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,

```

Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkie, P.S., Jr., Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M., Moreno-Palancas, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Saudke, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y., Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dinko, D., Feng, D.-F., Ferrite, A., Fischer, C., Hastings, G.A., He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Meissner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.

TITLE
Initial assessment of human gene diversity and expression patterns based upon 85 million nucleotides of cDNA sequence

JOURNAL
Nature 377 (6547 Suppl), 3-174 (1995)

MEDLINE
96026280

COMMENT

Other ESTs: THC168663

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlavet@tigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/tldb/hgi/hgi.html>)

Seq primer: M3 Reverse.

FEATURES
Location/Qualifiers

source

1..318

/organism="Homo sapiens"

/db_xref="ATCC (inhost):193847"

/db_xref="taxon:9606"

/clone_lib="Uterus"

/sex="female"

/dev_stage="adult"

/note="Organ: uterus; Vector: pbluescript SK-; Site_1:

ECORI; Site_2: XhoI"

BASE COUNT 96 a 66 c 72 g 83 t 1 others

ORIGIN

1..318

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="Uterus"

/sex="female"

/dev_stage="adult"

/note="Organ: uterus; Vector: pbluescript SK-; Site_1:

ECORI; Site_2: XhoI"

BASE COUNT 96 a 66 c 72 g 83 t 1 others

ORIGIN

1..318

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="Uterus"

/sex="female"

/dev_stage="adult"

/note="Organ: uterus; Vector: pbluescript SK-; Site_1:

ECORI; Site_2: XhoI"

BASE COUNT 96 a 66 c 72 g 83 t 1 others

ORIGIN

1..318

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="Uterus"

/sex="female"

/dev_stage="adult"

/note="Organ: uterus; Vector: pbluescript SK-; Site_1:

ECORI; Site_2: XhoI"

BASE COUNT 96 a 66 c 72 g 83 t 1 others

ORIGIN

1..318

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="Uterus"

/sex="female"

/dev_stage="adult"

/note="Organ: uterus; Vector: pbluescript SK-; Site_1:

ECORI; Site_2: XhoI"

BASE COUNT 96 a 66 c 72 g 83 t 1 others

ORIGIN

1..318

seq_name: gb_est25:AI800231

seq_documentation_block:

LOCUS AI800231 464 bp mRNA EST 13-DEC-1999
DEFINITION ti76c07.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2137932 3'
similar to TR:075556 075556 MAMMAGLOBIN B PRECURSOR. ; mRNA
sequence.

ACCESSION AI800231

VERSION AI800231.1 GI:5365703

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 464)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone Distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/dbp/image/image.html

Insert Length: 575 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 453.

Location/Qualifiers

1..464

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2137932"

/clone_lib="NCI_CGAP_Kid11"

/lab_host="DH10B"

/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with

a modified polylinker; Site_1: Not I; Site_2: Eco RI;

plasmid DNA from the normalized library NCI_CGAP_Kid3 was

prepared, and ss circles were made in vitro. Following HAP

purification, this DNA was used as tracer in a subtractive

hybridization reaction. The driver was PCR-amplified cDNAs

from a pool of 5,000 clones made from the same library

(cloneId: 1322376-1323911, 1456007-1456775, and

1500552-1502855). Subtraction by Bento Soares and M.

Fatima Bonaldo.

BASE COUNT 131 a 97 c 105 g 131 t

ORIGIN

1..131

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2137932"

/clone_lib="NCI_CGAP_Kid11"

/lab_host="DH10B"

/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with

a modified polylinker; Site_1: Not I; Site_2: Eco RI;

plasmid DNA from the normalized library NCI_CGAP_Kid3 was

prepared, and ss circles were made in vitro. Following HAP

purification, this DNA was used as tracer in a subtractive

hybridization reaction. The driver was PCR-amplified cDNAs

from a pool of 5,000 clones made from the same library

(cloneId: 1322376-1323911, 1456007-1456775, and

1500552-1502855). Subtraction by Bento Soares and M.

Fatima Bonaldo.

BASE COUNT 131 a 97 c 105 g 131 t

ORIGIN

1..131

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2137932"

/clone_lib="NCI_CGAP_Kid11"

/lab_host="DH10B"

/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with

a modified polylinker; Site_1: Not I; Site_2: Eco RI;

plasmid DNA from the normalized library NCI_CGAP_Kid3 was

prepared, and ss circles were made in vitro. Following HAP

purification, this DNA was used as tracer in a subtractive

hybridization reaction. The driver was PCR-amplified cDNAs

from a pool of 5,000 clones made from the same library

(cloneId: 1322376-1323911, 1456007-1456775, and

1500552-1502855). Subtraction by Bento Soares and M.

Fatima Bonaldo.

BASE COUNT 131 a 97 c 105 g 131 t

ORIGIN

1..131

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2137932"

/clone_lib="NCI_CGAP_Kid11"

/lab_host="DH10B"

/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with

a modified polylinker; Site_1: Not I; Site_2: Eco RI;

plasmid DNA from the normalized library NCI_CGAP_Kid3 was

prepared, and ss circles were made in vitro. Following HAP

purification, this DNA was used as tracer in a subtractive

hybridization reaction. The driver was PCR-amplified cDNAs

from a pool of 5,000 clones made from the same library

(cloneId: 1322376-1323911, 1456007-1456775, and

1500552-1502855). Subtraction by Bento Soares and M.

Fatima Bonaldo.

BASE COUNT 131 a 97 c 105 g 131 t

ORIGIN

1..131

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2137932"

/clone_lib="NCI_CGAP_Kid11"

/lab_host="DH10B"

/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with

a modified polylinker; Site_1: Not I; Site_2: Eco RI;

plasmid DNA from the normalized library NCI_CGAP_Kid3 was

prepared, and ss circles were made in vitro. Following HAP

purification, this DNA was used as tracer in a subtractive

hybridization reaction. The driver was PCR-amplified cDNAs

from a pool of 5,000 clones made from the same library

(cloneId: 1322376-1323911, 1456007-1456775, and

1500552-1502855). Subtraction by Bento Soares and M.

Fatima Bonaldo.

BASE COUNT 131 a 97 c 105 g 131 t

ORIGIN

1..131

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2137932"

/clone_lib="NCI_CGAP_Kid11"

/lab_host="DH10B"

/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with

a modified polylinker; Site_1: Not I; Site_2: Eco RI;

plasmid DNA from the normalized library NCI_CGAP_Kid3 was

prepared, and ss circles were made in vitro. Following HAP

purification, this DNA was used as tracer in a subtractive

hybridization reaction. The driver was PCR-amplified cDNAs

from a pool of 5,000 clones made from the same library

(cloneId: 1322376-1323911, 1456007-1456775, and

1500552-1502855). Subtraction by Bento Soares and M.

Fatima Bonaldo.

BASE COUNT 131 a 97 c 105 g 131 t

ORIGIN

1..131

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2137932"

/clone_lib="NCI_CGAP_Kid11"

/lab_host="DH10B"

/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with

a modified polylinker; Site_1: Not I; Site_2: Eco RI;

plasmid DNA from the normalized library NCI_CGAP_Kid3 was

prepared, and ss circles were made in vitro. Following HAP

purification, this DNA was used as tracer in a subtractive

hybridization reaction. The driver was PCR-amplified cDNAs

from a pool of 5,000 clones made from the same library

(cloneId: 1322376-1323911, 1456007-1456775, and

1500552-1502855). Subtraction by Bento Soares and M.

Fatima Bonaldo.

BASE COUNT 131 a 97 c 105 g 131 t

ORIGIN

1..131

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2137932"

/clone_lib="NCI_CGAP_Kid11"

/lab_host="DH10B"

/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with

a modified polylinker; Site_1: Not I; Site_2: Eco RI;

plasmid DNA from the normalized library NCI_CGAP_Kid3 was

prepared, and ss circles were made in vitro. Following HAP

purification, this DNA was used as tracer in a subtractive

hybridization reaction. The driver was PCR-amplified cDNAs

Tue May 1 11:47:50 2001

```
34 erAspAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
|||||
301 GTGATGCCGCTGCAGAGCTATGGGAAATTCAAAGCAGTGTTCCTCAAC 252
|||||
51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTy 67
|||||
251 CAGTCACATAGAACTCTGAAAAAATTGGGACTGATGATGCATACAGTGTA 202
|||||
67 rAspSerIleTrpCysAsnLeu 74
|||||
201 CGACAGCATTTGGTGTAAATATG 180
|||||
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 15:12:13 ; Search time 12.98 Seconds
(without alignments)
109.523 Million cell updates/sec

Title: US-09-367-009-3
Perfect score: 392
Sequence: 1 DSGCKLLEDMVEKTSINSIS.....TLKNFGLMHVYDSIWCNL 74

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2.6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2.6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2.6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2.6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2.6/ptodata/2/iaa/PTUS_COMB.pep.*
6: /cgn2.6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	390	99.5	95	3	US-08-821-451A-6
2	390	99.5	95	4	US-09-263-810-6
3	208	53.1	93	1	US-08-455-896-2
4	208	53.1	93	2	US-08-933-149-2
5	208	53.1	93	2	US-09-082-343-2
6	208	53.1	93	3	US-09-082-253-2
7	208	53.1	93	5	PCT-US96-08235-2
8	164	41.8	95	3	US-08-821-451A-27
9	164	41.8	95	4	US-09-263-810-27
10	163	41.6	95	1	US-08-455-896-7
11	163	41.6	95	2	US-08-933-149-7
12	163	41.6	95	2	US-09-082-343-7
13	163	41.6	95	3	US-09-082-253-7
14	163	41.6	95	5	PCT-US96-08235-7
15	61.5	15.7	1257	4	US-08-947-823-5
16	59.5	15.2	114	1	US-08-031-399-3
17	59.5	15.2	114	1	US-08-031-399-6
18	59.5	15.2	114	1	US-08-031-399-12
19	59.5	15.2	114	1	US-08-393-305-3
20	59.5	15.2	114	1	US-08-393-305-6
21	59.5	15.2	114	1	US-08-726-817-3
22	59.5	15.2	114	1	US-08-726-817-6
23	59.5	15.2	114	1	US-08-504-042-3
24	59.5	15.2	114	1	US-08-504-042-6
25	59.5	15.2	114	1	US-08-504-042-12
26	59.5	15.2	114	2	US-08-725-969-3
27	59.5	15.2	114	2	US-08-725-969-6

28	59.5	15.2	114	2	US-08-794-524-3	Sequence 3, Appli
29	59.5	15.2	114	2	US-08-794-524-6	Sequence 6, Appli
30	59.5	15.2	114	4	US-09-189-193-3	Sequence 3, Appli
31	59.5	15.2	114	4	US-09-189-193-6	Sequence 6, Appli
32	59.5	15.2	114	5	PCT-US94-03793-3	Sequence 3, Appli
33	59.5	15.2	114	5	PCT-US94-03793-6	Sequence 6, Appli
34	59.5	15.2	114	5	PCT-US94-03793-12	Sequence 12, Appli
35	59.5	15.2	122	1	US-08-300-903A-3	Sequence 3, Appli
36	59.5	15.2	122	1	US-08-031-399-2	Sequence 2, Appli
37	59.5	15.2	162	1	US-08-031-399-5	Sequence 5, Appli
38	59.5	15.2	162	1	US-08-393-305-2	Sequence 2, Appli
39	59.5	15.2	162	1	US-08-393-305-5	Sequence 5, Appli
40	59.5	15.2	162	1	US-08-535-733-2	Sequence 2, Appli
41	59.5	15.2	162	1	US-08-284-393B-9	Sequence 9, Appli
42	59.5	15.2	162	1	US-08-726-817-2	Sequence 2, Appli
43	59.5	15.2	162	1	US-08-726-817-5	Sequence 5, Appli
44	59.5	15.2	162	1	US-08-504-042-2	Sequence 2, Appli
45	59.5	15.2	162	1	US-08-504-042-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-08-821-451A-6
; Sequence 6, Application US/08821451A
; Patent No. 6056724
; GENERAL INFORMATION:
; APPLICANT: Jian Ni, Guo-Liang Yu and Reiner Gentz
; TITLE OF INVENTION: Human Endometrial Specific Steroid-
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/821,451A
FILING DATE: March 21, 1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/014,724
FILING DATE: March 21, 1996
ATTORNEY/AGENT INFORMATION:
NAME: MULLINS, J.G.
REGISTRATION NUMBER: 33,073
REFERENCE/DOCKET NUMBER: 325800-521 (PF257)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 95 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS:
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-08-821-451A-6

Query Match 99.5%; Score 390; DB 3; Length 95;
Best Local Similarity 98.6%; Pred. No. 5.3e-43;
Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DSGCKLEDVMEKTSINSDISPEYKELQEFIDSDAAAEAMGKFKQCFLNOSHRTLNKFG 60
Db 19 DSGCKLEDVMEKTSINSDISPEYKELQEFIDSDAAAEAMGKFKQCFLNOSHRTLNKFG 78
Oy 61 LMHTVYDSIWCNL 74
Db 79 LMHTVYDSIWCNM 92

RESULT 2

US-09-263-810-6
; Sequence 6, Application US/09263810
; Patent No. 6174992

GENERAL INFORMATION:

APPLICANT: JIAN NI, Guo-Liang Yu and Reiner Gentz
TITLE OF INVENTION: Human Endometrial Specific Steroid-
TITLE OF INVENTION: Binding Factor I, II and III
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:

ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2

OPERATING SYSTEM: MS-DOS

SOFTWARE: WORD PERFECT 5.1

APPLICATION DATA:

APPLICATION NUMBER: US/09/263,810
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/821,451

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: MULLINS, J.G.

REGISTRATION NUMBER: 33,073

REFERENCE/DOCKET NUMBER: 325800-521 (PF257)

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-994-1700

TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 95 AMINO ACIDS

TYPE: AMINO ACID

STRANDEDNESS:

TOPOLOGY: LINEAR

MOLECULE TYPE: PROTEIN

US-09-263-810-6

Query Match

Best Local Similarity 99.5%; Score 390; DB 4; Length 95;

Patent No. 53e-43;

Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DSGCKLEDVMEKTSINSDISPEYKELQEFIDSDAAAEAMGKFKQCFLNOSHRTLNKFG 60
Db 19 DSGCKLEDVMEKTSINSDISPEYKELQEFIDSDAAAEAMGKFKQCFLNOSHRTLNKFG 78
Oy 61 LMHTVYDSIWCNL 74
Db 79 LMHTVYDSIWCNM 92

RESULT 3

US-08-455-896-2

; Sequence 2, Application US/08455896

; Patent No. 5668267

GENERAL INFORMATION:

APPLICANT: WATSON, MARK A.
APPLICANT: FLEMING, TIMOTHY P.
TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:

ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: USA
ZIP: 63105-1817

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/455,896

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: HOLLAND, DONALD R.

REGISTRATION NUMBER: 35,197

REFERENCE/DOCKET NUMBER: 952726

TELECOMMUNICATION INFORMATION:

TELEPHONE: (314) 727-5188

TELEFAX: (314) 727-6092

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 93 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

US-08-455-896-2

Query Match

Best Local Similarity 53.1%; Score 208; DB 1; Length 93;

Patent No. 1.2e-19;

Matches 39; Conservative 12; Mismatches 22; Indels 0; Gaps 0;

Oy 2 SGCPLLENVISKTNPOVSKTEYKELQEFIDSDAAAEAMGKFKQCFLNOSHRTLNKFL 61
Db 20 SGCPLLENVISKTNPOVSKTEYKELQEFIDSDAAAEAMGKFKQCFLNOSHRTLNKFL 79
Oy 62 MMHTVYDSIWCNL 74
Db 80 FMOLIYDSSLCDL 92

RESULT 4

US-08-933-149-2

; Sequence 2, Application US/08933149

; Patent No. 5922836

GENERAL INFORMATION:

APPLICANT: WATSON, MARK A.

APPLICANT: FLEMING, TIMOTHY P.

TITLE OF INVENTION: MAMMAGLOBIN, A SECRETED

TITLE OF INVENTION: MAMMARY SPECIFIC BREAST CANCER PROTEIN

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: HOWELL & HAFERKAMP, L.C.

STREET: 7733 FORSYTH BOULEVARD, SUITE 1400

CITY: ST. LOUIS

STATE: MISSOURI

COUNTRY: USA

ZIP: 63105-1817

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS